Introductory Physics I

Elementary Mechanics and Applications

by

Robert G. Brown

Duke University Physics Department
Durham, NC 27708-0305
rgb@phy.duke.edu

Contents

Preface

This ***introductory mechanics*** text is intended to be used in the first semester of a two-semester series of courses teaching ***introductory physics*** at the college level, followed by a second semester course in ***introductory electricity and magnetism, and optics***. The text is intended to support teaching the material at a rapid, but ***advanced*** level – it was developed to support teaching introductory calculus-based physics to potential physics majors, engineers, and other natural science majors at Duke University over a period of more than thirty years.

Students who hope to succeed in learning physics from this text will need, as a minimum prerequisite, a ***solid grasp of basic mathematics***. It is strongly recommended that all students have mastered mathematics at least through single-variable differential calculus (typified by the AB advanced placement test or a first-semester college calculus course). Students should also be *taking* (or have completed) single variable integral calculus (typified by the BC advanced placement test or a second-semester college calculus course). In the text it is presumed that students are competent in geometry, trigonometry, algebra, and single variable calculus; more advanced multivariate calculus is used in a number of places but it is taught in context as it is needed and is always "separable" into two or three independent one-dimensional integrals.

Many students are, unfortunately *weak* in their mastery of mathematics at the time they take physics. This enormously complicates the process of learning for them, especially if they are years removed from when they took their algebra, trig, and calculus classes (as is frequently the case for pre-medical students taking the course in their junior year of college). For that reason, a separate supplementary text intended ***specifically to help students of introductory physics quickly and efficiently review the required math*** is being prepared as a companion volume to all semesters of introductory physics. Indeed, it should really be quite useful for any course being taught with any textbook series and not just this one.

This book is located here:

http://www.phy.duke.edu/~rgb/Class/math_for_intro_physics.php

and I *strongly suggest* that all students who are reading these words preparing to begin studying physics pause for a moment, visit this site, and either download the pdf or bookmark the site.

Note that *Week 0: How to Learn Physics* is not part of the course *per se*, but I usually do a quick review of this material (as well as the course structure, grading scheme, and so on) in my first lecture of any given semester, the one where students are still finding the room, dropping and adding courses, and one cannot present real content in good conscience unless you plan to do it again in the second lecture as well. Students *greatly benefit* from guidance on how to study, as most enter physics thinking that they can master it with nothing but the memorization and rote learning skills that have served them so well for their many other fact-based classes. Of course this is completely false – physics is *reason* based and *conceptual* and it requires a very different pattern of study than simply staring at and trying to memorize lists of formulae or examples.

Students, however, should not count on their instructor doing this – they need to be self-actualized in their study from the beginning. It is therefore *strongly suggested* that all students read this preliminary chapter right away as their first "assignment" whether or not it is covered in the first lecture or assigned. In fact, (if you're just such a student reading these words) you can always decide to read it *right now* (as soon as you finish this Preface). It won't take you an hour, and might make as much as a full letter difference (to the good) in your final grade. What do you have to lose?

Even if you think that you are an excellent student and learn things totally effortlessly, I strongly suggest reading it. It describes a new perspective on the teaching and learning process supported by very recent research in neuroscience and psychology, and makes very specific suggestions as to the best way to proceed to learn physics.

Finally, the *Introduction* is a rapid summary of *the entire course!* If you read it and look at the pictures *before* beginning the course proper you can get a good conceptual overview of everything you're going to learn. If you *begin* by learning in a *quick* pass the broad strokes for the whole course, when you go through each chapter in all of its detail all those facts and ideas have a place to live in your mind.

That's the primary idea behind this textbook – in order to be easy to remember, ideas need a house, a place to live. Most courses try to build you that house by giving you one nail and piece of wood at a time, and force you to build it in complete detail from the ground up.

Real houses aren't built that way at all! First a foundation is established, then the *frame of the whole house* is erected, and then, slowly but surely, the frame is wired and plumbed and drywalled and finished with all of those picky little details. It works better that way. So it is with learning.

Textbook Layout and Design

This textbook has a design that is just about perfectly backwards compared to most textbooks that currently cover the subject. Here are its primary design features:

- All mathematics required by the student is reviewed in a standalone, cross-referenced (free) work at the *beginning* of the book rather than in an appendix that many students never find.
- There are only *twelve chapters*. The book is organized so that it can be sanely taught in a *single college semester* with at *most* a chapter a week.
- It *begins* each chapter with an "abstract" and chapter summary. Detail, especially lecture-note style mathematical detail, follows the summary rather than the other way around.

- This text does *not* spend page after page trying to explain in English how physics works (prose which to my experience nobody reads anyway). Instead, a terse "lecture note" style presentation outlines the main points and presents considerable mathematical detail to support solving problems.

- Verbal and conceptual understanding *is*, of course, very important. It is expected to come from verbal instruction and discussion in the classroom and recitation and lab. This textbook *relies* on having a committed and competent instructor and a sensible learning process.

- Each chapter ends with a *short* (by modern standards) selection of *challenging* homework problems. A good student might well get through all of the problems in the book, rather than at most 10% of them as is the general rule for other texts.

- The problems are weakly sorted out by level, as this text is intended to support non-physics science and pre-health profession students, engineers, and physics majors all three. The *material* covered is of course the same for all three, but the level of detail and difficulty of the math used and required is a bit different.

- The textbook is entirely algebraic in its presentation and problem solving requirements – with *very few exceptions* no calculators should be required to solve problems. The author assumes that any student taking physics is capable of punching numbers into a calculator, but it is *algebra* that ultimately determines the formula that they should be computing. Numbers are used in problems only to illustrate what "reasonable" numbers might be for a given real-world physical situation or where the problems cannot reasonably be solved algebraically (e.g. resistance networks).

This layout provides considerable benefits to both instructor and student. This textbook supports a *top-down* style of learning, where one learns each distinct chapter topic by quickly getting the main points onboard via the summary, then derives them or explores them in detail, then applies them to example problems. Finally one uses what one has started to learn working in groups and with direct mentoring and support from the instructors, to solve highly challenging problems that *cannot* be solved without acquiring the deeper level of understanding that is, or should be, the goal one is striving for.

It's without doubt a lot of work. Nobody said learning physics would be *easy*, and this book certainly doesn't claim to make it so. However, this approach will (for most students) *work*.

The reward, in the end, is the ability to see the entire world around you through new eyes, understanding much of the "magic" of the causal chain of physical forces that makes all things unfold in time. Natural Law is a strange, beautiful sort of magic; one that is utterly impersonal and mechanical and yet filled with structure and mathematics and light. It *makes sense*, both in and of itself and of the physical world you observe.

Enjoy.

I: Getting Ready to Learn Physics

Preliminaries

1.1: See, Do, *Teach*

If you are reading this, I assume that you are either taking a course in physics or wish to learn physics on your own. If this is the case, I want to begin by teaching you the importance of your personal *engagement* in the learning process. If it comes right down to it, how well you learn physics, how good a grade you get, and how much *fun* you have all depend on how enthusiastically you tackle the learning process. If you remain disengaged, detatched from the learning process, you almost certainly will do poorly and be miserable while doing it. If you can find *any degree* of engagement – or open enthusiasm – with the learning process you will very likely do well, or at least as well as possible.

Note that I use the term *learning*, not *teaching* – this is to emphasize from the beginning that learning is a choice and that *you* are in control. Learning is active; being taught is passive. It is up to you to *seize control* of your own educational process and *fully participate*, not sit back and wait for knowledge to be forcibly injected into your brain.

You may find yourself stuck in a course that is taught in a traditional way, by an instructor that lectures, assigns some readings, and maybe on a good day puts on a little dog-and-pony show in the classroom with some audiovisual aids or some demonstrations. The standard expectation in this class is to sit in your chair and watch, passive, taking notes. No real engagement is "required" by the instructor, and lacking activities or a structure that encourages it, you lapse into becoming a lecture transcription machine, recording all kinds of things that make no immediate sense to you and telling yourself that you'll sort it all out later.

You may find yourself floundering in such a class – for good reason. The instructor presents an ocean of material in each lecture, and you're going to actually retain at most a few cupfuls of it functioning as a scribe and passively copying his pictures and symbols without first extracting their sense. And the lecture *makes* little sense, at least at first, and reading (if you do any reading at all) does little to help. Demonstrations can sometimes make one or two ideas come clear, but only at the expense of twenty other things that the instructor now has no time to cover and expects you to get from the readings alone. You continually postpone going over the lectures and readings to understand the material any more than is strictly required to do the homework, until one day a *big test* draws nigh and you realize that you really don't understand anything and have forgotten most of what you did, briefly, understand. Doom and destruction loom.

Sound familiar?

On the other hand, you may be in a course where the instructor has structured the course with a balanced mix of *open* lecture (held as a freeform discussion where questions aren't just encouraged but required) and group interactive learning situations such as a carefully structured recitation and lab where discussion and doing blend together, where students teach each other and use what they have learned in many ways and contexts. If so, you're lucky, but luck only goes so far.

Even in a course like this you may *still* be floundering because you may not understand *why* it is important for you to participate with your whole spirit in the quest to learn anything you ever choose to study. In a word, you simply may not give a rodent's furry behind about learning the material so that studying is always a fight with yourself to "make" yourself do it – so that no matter what happens, *you lose*. This too may sound very familiar to some.

The importance of engagement and participation in "active learning" (as opposed to passively being taught) is not really a new idea. Medical schools were four year programs in the year 1900. They are four year programs today, where the amount of information that a physician must now master in those four years is probably *ten times greater* today than it was back then. Medical students are necessarily among the most efficient learners on earth, or they simply cannot survive.

In medical schools, the optimal learning strategy is compressed to a three-step adage: See one, do one, teach one.

See a procedure (done by a trained expert).

Do the procedure yourself, with the direct supervision and guidance of a trained expert.

Teach a student to do the procedure.

See, do, teach. Now you *are* a trained expert (of sorts), or at least so we devoutly hope, because that's all the training you are likely to get until you start doing the procedure over and over again with real humans and with limited oversight from an attending physician with too many other things to do. So you practice and study on your own until you achieve real mastery, because a mistake can *kill* somebody.

This recipe is quite general, and can be used to increase *your own* learning in almost *any* class. In fact, lifelong success in learning with or without the guidance of a good teacher is a matter of discovering the importance of *active engagement and participation* that this recipe (non-uniquely) encodes. Let us rank learning methodologies in terms of "probable degree of active engagement of the student". By probable I mean the degree of active engagement that I as an instructor have observed in students over many years and which is significantly reinforced by research in teaching methodology, especially in physics and mathematics.

Listening to a lecture as a transcription machine with your brain in "copy machine" mode is almost entirely passive and is for *most* students *probably* a nearly complete waste of time. That's not to say that "lecture" in the form of an organized presentation and review of the material to be learned isn't important or is completely useless! It serves one *very important purpose* in the grand scheme of learning, but by being passive *during* lecture *you* cause it to fail in its purpose. Its purpose is *not* to give you a complete, line by line transcription of the words of your instructor to ponder later and alone. It is to convey, for a brief shining moment, the *sense* of the *concepts* so that you *understand them*.

It is difficult to sufficiently emphasize this point. If lecture doesn't make sense *to you* when the instructor presents it, you will have to work much harder to achieve the sense of the material "later", if later ever comes at all. If you fail to identify the important concepts during the presentation and see the lecture as a string of disconnected facts, you will have to remember *each* fact as if it were an abstract string of symbols, placing impossible demands on your memory even if you are extraordinarily bright. If you fail to achieve some degree of understanding (or *synthesis* of the material, if you prefer) in lecture by asking questions and getting expert explanations on the spot, you will have to build it later out of your notes on a set of abstract symbols that made no sense to you at the time. You might as well be trying to translate Egyptian Hieroglyphs without a Rosetta Stone, and the best of luck to you with *that*.

Reading is a bit more active – at the very least your brain is more likely to be somewhat engaged if you aren't "just" transcribing the book onto a piece of paper or letting the words and symbols happen in your mind – but is still pretty passive. Even watching nifty movies or cool-ee-oh demonstrations is basically sedentary – you're still just sitting there while somebody or something *else* makes it all happen in your brain while you aren't *doing* much of anything. At best it grabs your attention a bit better (on average) than lecture, but *you* are mentally *passive*.

In all of these forms of learning, the single active thing you are likely to be doing is taking notes or moving an eye muscle from time to time. For better or worse, the human brain isn't designed to learn well in passive mode. Parts of your brain are likely to take charge and pull your eyes irresistably to the window to look outside where *active* things are going on, things that might not be so damn *boring!*

With your active engagement, with your taking charge of and participating in the learning process, things change dramatically. Instead of passively listening in lecture, you can at least *try* to ask questions and initiate discussions whenever an idea is presented that makes no intial sense to you. Discussion is an *active* process even if you aren't the one talking at the time. *You participate!* Even a tiny bit of participation in a classroom setting where students are constantly asking questions, where the instructor is constantly answering them and asking the students questions in turn makes a huge difference. Humans being social creatures, it also makes the class a lot more fun!

In summary, sitting on your ass[1] and writing meaningless (to you, so far) things down as somebody says them in the hopes of being able to "study" them and discover their meaning on your own later is *boring* and for most students, later never comes because you are busy with *many* classes, because you haven't discovered anything beautiful or exciting (which is the *reward* for figuring it all out – if you ever get there) and then there is partying and hanging out with friends and having *fun*. Even if you do find the time and really want to succeed, in a complicated subject like physics you are less likely to be *able* to discover the meaning on your own (unless you are *so bright* that learning methodology is irrelevant and you learn in a single pass no matter what). Most introductory students are swamped by the details, and have small chance of discovering the *patterns* within those details that constitute "making sense" and make the detailed information *much, much easier to learn* by enabling a compression of the detail into a much smaller set of connected ideas.

Articulation of ideas, whether it is to yourself or to others in a discussion setting, *requires* you to create tentative patterns that might describe and organize all the details you are being

[1] I mean, of course, your donkey. What did you think I meant?

presented with. Using those patterns and applying them to the details as they are presented, you naturally encounter places where your tentative patterns are wrong, or don't quite work, where something "doesn't make sense". In an "active" lecture students participate in the process, and can ask questions and kick ideas around until they *do* make sense. Participation is also *fun* and helps you pay far more attention to what's going on than when you are in passive mode. It may be that this increased attention, this consideration of many alternatives and rejecting some while retaining others with social reinforcement, is what makes all the difference. To learn optimally, even "seeing" must be an active process, one where you are not a vessel waiting to be filled through your eyes but rather part of a team studying a puzzle and looking for the patterns *together* that will help you eventually solve it.

Learning is increased still further by *doing*, the very essence of activity and engagement. "Doing" varies from course to course, depending on just what there is for you to do, but it always is the *application* of what you are learning to some sort of activity, exercise, problem. It is *not* just a recapitulation of symbols: "looking over your notes" or "(re)reading the text". The symbols for any given course of study (in a physics class, they very likely will *be* algebraic symbols for real although I'm speaking more generally here) do not, initially, mean a lot to you. If I write $\vec{F} = q(\vec{v} \times \vec{B})$ on the board, it means a great deal to *me*, but if you are taking this course for the first time it probably means zilch to *you*, and yet I pop it up there, draw some pictures, make some noises that hopefully make sense to you at the time, and blow on by. Later you read it in your notes to try to recreate that sense, but you've *forgotten* most of it. Am I describing the income I expect to make selling $\vec{B}$ tons of barley with a market value of $\vec{v}$ and a profit margin of q?

To *learn* this expression (for yes, this is a force law of nature and one that we very much must learn this semester) we have to learn what the symbols stand for – q is the charge of a point-like object in motion at velocity $\vec{v}$ in a magnetic field $\vec{B}$, and $\vec{F}$ is the resulting force acting on the particle. We have to learn that the $\times$ symbol is the *cross product of evil* (to most students at any rate, at least at first). In order to get a *gut feeling* for what this equation represents, for the directions associated with the cross product, for the trajectories it implies for charged particles moving in a magnetic field in a variety of contexts one has to *use* this expression to solve problems, *see* this expression in action in laboratory experiments that let you prove to yourself that it isn't bullshit and that the world really does have cross product force laws in it. You have to do your homework that involves this law, and be fully engaged.

The learning process isn't exactly linear, so if you participate fully in the discussion and the doing while going to even the most traditional of lectures, you have an excellent chance of getting to the point where you can score anywhere from a 75% to an 85% in the course. In most schools, say a C+ to B+ performance. Not bad, but not really excellent. A few students will still get A's – they either work extra hard, or really like the subject, or they have some sort of secret, some way of getting over that barrier at the 90's that is only crossed by those that really do understand the material quite well.

Here is the secret for getting *yourself* over that 90% hump, even in a physics class (arguably one of the most difficult courses you can take in college), even if you're *not* a super-genius (or have never managed in the past to learn like one, a glance and you're done): *Work in groups!*

That's it. Nothing really complex or horrible, just get together with your friends who are also taking the course and do your homework *together*. In a well designed physics course (and many courses in mathematics, economics, and other subjects these days) you'll have *some* aspects of the class, such as a recitation or lab, where you are *required* to work in groups, and the groups and group activities may be highly structured or freeform. "Studio" or "Team Based Learning" methods for teaching physics have even wrapped the lecture itself into a group-structured setting, so *everything* is done in groups/teams, and (probably by making it nearly impossible to be disengaged and sit passively in class waiting for learning to "happen") this approach yields measureable improvements (all things being equal) on at least some objective instruments for measurement of learning.

If you take charge of your own learning, though, you will quickly see that in *any* course, however taught, *you can study in a group!* This is true even in a course where "the homework]" is to be done alone by fiat of the (unfortunately ignorant and misguided) instructor. Just study "around" the actual assignment – assign *yourselves* problems "like" the actual assignment – most textbooks have plenty of extra problems and then there is the Internet and other textbooks – and do them in a group, then (afterwards!) break up and do your actual assignment alone. Note that if you use a completely different textbook to pick your group problems from and do them together before *looking* at your assignment in *your* textbook, you can't even be blamed if some of the ones you pick turn out to be ones your instructor happened to assign.

Oh, and not-so-subtly – give the instructor a PDF copy of this book (it's free for instructors, after all, and a click away on the Internet) and point to this page and paragraph containing the following little message from me to them:

> Yo! Teacher! Let's wake up and smell the coffee! Don't prevent your students from doing homework in groups – require it! Make the homework correspondingly more difficult! Give them quite a lot of course credit for doing it well! Construct a recitation or review session where students – in groups – who still cannot get the most difficult problems can get socratic tutorial help *after* working hard on the problems on their own! Integrate discussion and deliberately teach to increase *active engagement* (instead of passive wandering attention) in lecture[2]. Then watch as student performance and engagement spirals into the stratosphere compared to what it was before...

Then pray. Some instructors have their egos tied up in things to the point where *they* cannot learn, and then what can you do? If an instructor lets ego or politics obstruct their search for functional methodology, you're screwed anyway, and you might as well just tackle the material on your own. Or heck, maybe their expertise and teaching experience vastly exceeds my own so that their naked words *are* sufficiently golden that any student should be able to learn by just hearing them and doing homework all alone in isolation from any peer-interaction process that might be of use to help them make sense of it all – all data to the contrary.

---

[2]Perhaps by using Team Based Learning methods to structure and balance student groups and "flipping" classrooms to foist the lecture off onto videos of somebody else lecturing to increase the time spent in the class working in groups, but I've found that in mid-sized classes and smaller (less than around fifty students) one can get very good results from traditional lecture without a specially designed classroom by the Chocolate Method – I lecture without notes and offer a piece of chocolate or cheap toy or nifty pencil to any student who catches me making a mistake on the board before I catch it myself, who asks a particularly good question, who looks like they are nodding off to sleep (seriously, chocolate works wonders here, especially when ceremoniously offered). Anything that keeps students *focussed* during lecture by making it into a game, by allowing/encouraging them to speak out without raising their hands, by praising them and rewarding them for engagement makes a huge difference.

My own words and lecture – in spite of my 31 years of experience in the classroom, in spite of the fact that it has been well over twenty years since I actually used lecture notes to teach the course, in spite of the fact I never, ever prepare for recitation because solving the homework problems with the students "cold" *as* a peer member of their groups is useful where copying my privately worked out solutions onto a blackboard for them to passively copy on their papers in turn is useless, in spite of the fact that I wrote this book *similarly* without the use of any outside resource – my words and lecture are *not*. On the other hand, students who work effectively in groups and learn to use this book (and other resources) and do all of the homework "to perfection" might well learn physics quite well without my involvement at all!

Let's understand *why* working in groups has such a dramatic effect on learning. What happens in a group? Well, a lot of *discussion* happens, because humans working on a common problem like to talk. There is plenty of *doing* going on, presuming that the group has a common task list to work through, like a small mountain of really difficult problems that nobody can possibly solve working on their own and are *barely* within their abilities working as a group backed up by the course instructor! Finally, in a group everybody has the opportunity to *teach!*

The importance of teaching – not only seeing the lecture presentation with your whole brain actively engaged and participating in an ongoing discussion so that it makes sense at the time, not only doing lots of homework problems and exercises that apply the material in some way, but *articulating* what you have discovered in this process and *answering questions* that force you to consider and reject alternative solutions or pathways (or not) cannot be overemphasized. Teaching each other in a peer setting (ideally with mentorship and oversight to keep you from teaching each other *mistakes*) is *essential!*

This problem you "get", and teach *others* (and actually learn it better from teaching it than they do from your presentation – never begrudge the effort required to teach your group peers even if some of them are very slow to understand). The next problem you don't get but some *other* group member does – they get to teach *you*. In the end you all learn *far more* about every problem as a consequence of the struggle, the exploration of false paths, the discovery and articulation of the correct path, the process of discussion, resolution and agreement in teaching whereby *everybody* in the group reaches full understanding.

I would assert that it is all but *impossible* for someone to become a (halfway decent) teacher of *anything* without learning along the way that the absolute best way to learn *any* set of material deeply is to *teach* it – it is the very foundation of Academe and has been for two or three thousand years. It is, as we have noted, built right into the intensive learning process of medical school and graduate school in general. For some reason, however, we don't incorporate a teaching component in most *undergraduate* classes, which is a shame, and it is basically nonexistent in nearly all K-12 schools, which is an open tragedy.

As an engaged student *you don't have to live with that!* Put it there yourself, by incorporating group study and mutual teaching into your learning process *with or without the help or permission of your teachers!* A really smart and effective group soon learns to *iterate* the teaching – I teach you, and to make sure you got it you *immediately* use the material I taught you and try to articulate it back to me. Eventually everybody in the group understands, everybody in the group benefits, *everybody in the group gets the best possible grade on the material*. This process will actually make you (quite literally) more intelligent. You may or may not become smart enough to lock down an A, but you will get the best grade you are capable of getting, for your given investment of effort.

This is close to the ultimate in engagement – highly active learning, with all cylinders of your brain firing away on the process. You can *see* why learning is enhanced. It is simply a bonus, a sign of a just and caring God, that it is also a lot more *fun* to work in a group, especially in a relaxed context with food and drink present. Yes, I'm encouraging you to have "physics study parties" (or history study parties, or psychology study parties). Hold contests. Give silly prizes. See. Do. Teach.

1.2: Other Conditions for Learning

Learning isn't *only* dependent on the engagement pattern implicit in the See, Do, Teach rule. Let's absorb a few more True Facts about learning, in particular let's come up with a handful of things that can act as "switches" and turn your ability to learn on and off quite independent of how your instructor structures your courses. Most of these things aren't *binary* switches – they are more like dimmer switches that can be slid up between dim (but not off) and bright (but not fully on). Some of these switches, or environmental parameters, act together more powerfully than they act alone. We'll start with the most important pair, a pair that research has shown work together to potentiate or block learning.

Instead of just telling you what they are, arguing that they are important for a paragraph or six, and moving on, I'm going to give you an early opportunity to *practice* active learning in the context of reading a chapter on active learning. That is, I want you to participate in a tiny mini-experiment. It works a little bit better if it is done verbally in a one-on-one meeting, but it should still work well enough even if it is done in this text that you are reading.

I'm going to give you a string of ten or so digits and ask you to glance at it one time for a count of three and then look away. No fair peeking once your three seconds are up! Then I want you to do something else for at least a minute – anything else that uses your whole attention and interrupts your ability to rehearse the numbers in your mind in the way that you've doubtless learned permits you to learn other strings of digits, such as holding your mind blank, thinking of the phone numbers of friends or your social security number. Even rereading this paragraph will do.

At the end of the minute, try to recall the number I gave you and write down what you remember. Then turn back to right here and compare what you wrote down with the actual number.

Ready? (No peeking yet...) Set? Go!

Ok, here it is, in a footnote at the bottom of the page to keep your eye from naturally reading ahead to catch a glimpse of it while reading the instructions above[3].

How did you do?

If you are like most people, this string of numbers is a bit too long to get into your immediate memory or visual memory in only three seconds. There was very little time for rehearsal, and then you went and did something else for a bit right away that was supposed to *keep* you

[3] 1357986420 (one, two, three, quit and do something else for one minute...)

from rehearsing whatever of the string you *did* manage to verbalize in three seconds. Most people will get anywhere from the first three to as many as seven or eight of the digits right, but probably not in the correct order, unless...

...they are particularly smart or lucky and in that brief three second glance have time to notice that the number consists of all the digits used exactly once! Folks that happened to "see" this at a glance probably did better than average, getting all of the correct digits but maybe in not quite the correct order.

People who are downright *brilliant* (and equally lucky) realized in only three seconds (without cheating an extra second or three, you know who you are) that it consisted of the string of odd digits in ascending order followed by the even digits in descending order. Those people probably got it *all perfectly right* even without time to rehearse and "memorize" the string! Look again at the string, see the pattern now?

The moral of this little mini-demonstration is that it is *easy* to overwhelm the mind's capacity for processing and remembering "meaningless" or "random" information. A string of ten measely (apparently) random digits is too much to remember for one lousy minute, especially if you aren't given time to do rehearsal and all of the other things we have to make ourselves do to "memorize" meaningless information.

Of course things *changed radically* the instant I pointed out the pattern! At this point you could very likely go away and come back to this point in the text *tomorrow* or even *a year from now* and have an *excellent* chance of remembering this particular digit string, because it *makes sense* of a sort, and there are plenty of cues in the text to trigger recall of the particular pattern that "compresses and encodes" the actual string. You don't have to remember *ten* random things at all – only two and a half – odd ascending digits followed by the opposite (of both). Patterns rock!

This example has obvious connections to lecture and class time, and is one reason retention from lecture is so lousy. For *most* students, lecture in any nontrivial college-level course is a long-running litany of stuff they don't know yet. Since it is all new to them, it might as well be random digits as far as their cognitive abilities are concerned, at least at first. Sure, there is pattern there, but you have to *discover* the pattern, which requires *time* and a certain amount of *meditation* on all of the information. Basically, you have to have a chance for the pattern to jump out of the stream of information and punch the switch of the damn light bulb we all carry around inside our heads, the one that is endlessly portrayed in cartoons. That light bulb is *real* – it actually exists, in more than just a metaphorical sense – and if you study long enough and hard enough to obtain a sudden, epiphinaic realization in any topic you are studying, however trivial or complex (like the pattern exposed above) it is quite likely to be accompanied by a purely mental flash of "light". You'll know it when it happens to you, in other words, and it feels *great*.

Unfortunately, the instructor doesn't usually give students a *chance* to experience this in lecture. No sooner is one seemingly random factoid laid out on the table than along comes a new, apparently disconnected one that pushes it out of place long before we can either memorize it the hard way or make sense out of it so we can remember it with a lot less work. This isn't really anybody's fault, of course; the light bulb is quite unlikely to go off in lecture *just* from lecture no matter *what* you or the lecturer do – it is something that happens to the prepared mind at the end of a process, not something that just fires away every time you hear a new idea.

The humble and unsurprising conclusion I want you to draw from this silly little mini-experiment is that *things are easier to learn when they make sense!* A *lot* easier. In fact, things that don't make sense to you are never "learned" – they are at best memorized. Information can almost always be *compressed* when you discover the patterns that run through it, especially when the patterns all fit together into the marvelously complex and beautiful and mysterious process we call "deep understanding" of some subject.

There is one more example I like to use to illustrate how important this information compression is to memory and intelligence. I play chess, badly. That is, I know the legal moves of the game, and have no idea at all how to use them effectively to improve my position and eventually win. Ten moves into a typical chess game I can't recall how I got myself into the mess I'm typically in, and at the end of the game I probably can't remember *any* of what went on except that I got trounced, again.

A chess *master*, on the other hand, can play umpty games at once, blindfolded, against pitiful fools like myself and when they've finished winning them all they can go back and reconstruct *each one* move by move, criticizing each move as they go. Often they can remember the games in their entirety days or even years later.

This isn't just because they are *smarter* – *they* might be completely unable to derive the Lorentz group from first principles, and I can, and this doesn't automatically make me smarter than them either. It is because chess makes *sense* to them – they've achieved a deep understanding of the game, as it were – and they've built a complex meta-structure memory in their brains into which they can poke chess moves so that they can be retrieved extremely efficiently. This gives them the *attendant* capability of searching vast portions of the game tree at a glance, where I have to tediously work through each branch, one step at a time, usually omitting some really important possibility because I don't realize that that knight on the far side of the board can affect things on this side where we are both moving pieces.

This sort of "deep" (synthetic) understanding of physics is very much the goal of *this* course (the one in the textbook you are reading, since I use this intro in many textbooks), and to achieve it you must *not* memorize things as if they are random factoids, you must work to abstract the beautiful intertwining of patterns that compress all of those apparently random factoids into things that you can easily remember offhand, that you can easily reconstruct from the pattern even if you forget the details, and that you can search through at a glance. But the process I describe can be applied to learning pretty much anything, as patterns and structure exist in abundance in *all* subjects of interest. There are even sensible rules that govern or describe the anti-pattern of *pure randomness!*

There's one more important thing you can learn from thinking over the digit experiment. *Some* of you reading this very likely didn't do what I asked, you didn't play along with the game. Perhaps it was too much of a bother – you didn't want to waste a *whole minute* learning something by actually *doing* it, just wanted to read the damn chapter and get it over with so you could do, well, whatever the hell else it is you were planning to do today that's more important to you than physics or learning in other courses.

If you're one of these people, you probably don't remember *any* of the digit string at this point from actually seeing it – you never even *tried* to memorize it. A very few of you may actually be so terribly jaded that you don't even remember the little mnemonic *formula* I gave above for the digit string (although frankly, people that are *that* disengaged are probably not about to do things like actually read a textbook in the first place, so possibly not). After all, either way the string is pretty damn meaningless, pattern or not.

Pattern and meaning aren't exactly the same thing. There are all sorts of patterns one can find in random number strings, they just aren't "real" (where we could wax poetic at this point about information entropy and randomness and monkeys typing Shakespeare if this were a different course). So why bother wasting brain energy on even the *easy* way to remember this string when doing so is utterly unimportant to you in the grand scheme of all things?

From this we can learn the *second* humble and unsurprising conclusion I want you to draw from this one elementary thought experiment. *Things are easier to learn when you care about learning them!* In fact, they are damn near impossible to learn if you really *don't* care about learning them.

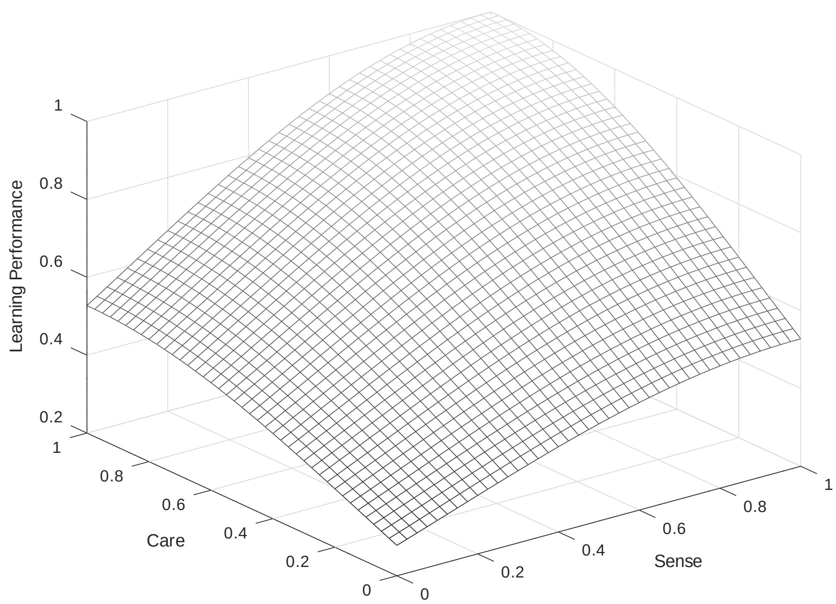


Figure 0.1: Conceptual relation between sense, care and learning, on simple relative scales.

Let's put the two observations together and plot them as a graph, just for fun (and because graphs help one learn for reasons we will explore just a bit in a minute). If you care about learning what you are studying, and the information you are trying to learn makes sense (if only for a moment, perhaps during lecture), the chances of your learning it are quite good. This alone isn't *enough* to guarantee that you'll learn it, but it they are basically both necessary conditions, and one of them is directly connected to degree of engagement.

On the other hand, if you care but the information you want to learn makes no sense, or if it makes sense but you hate the subject, the instructor, your school, your life and just don't care, your chances of learning it aren't so good, probably a bit better in the first case than in the second as if you care you have a *chance* of finding someone or some way that will help you make sense of whatever it is you wish to learn, where the person who doesn't cares, well, they don't care. Why should they remember it?

If you don't give a rat's ass about the material *and* it makes no sense to you, go home. Leave school. Do something else. You basically have almost no chance of learning the material unless you are gifted with a transcendent intelligence (wasted on a dilettante who lives in a state of perpetual ennui) and are miraculously gifted with the ability to learn things effortlessly even when they make no sense to you and you don't really care about them. All the learning tricks and study patterns in the world won't help a student who doesn't try, doesn't care, and for whom the material never makes sense.

If we worked at it, we could probably find other "logistic" controlling parameters to associate with learning – things that increase your probability of learning monotonically as they vary. Some of them are already apparent from the discussion above. Let's list a few more of them with explanations just so that you can see how *easy* it is to sit down to study and try to learn and have "something wrong" that decreases your ability to learn in that particular place and time.

Learning is actual work and involves a fair bit of biological stress, just like working out. Your brain needs *food* – it burns a whopping *20-30% of your daily calorie intake all by itself* just living day to day, even more when you are really using it or are somewhat sedentary in your physical habits. Note that your brain runs on pure, energy-rich glucose, so when your blood sugar drops your brain activity drops right along with it. This can happen (paradoxically) because you *just ate a carbohydrate rich meal*. A balanced diet containing foods with a lower glycemic index[4] ends to be harder to digest and provides a longer period of sustained energy for your brain. A daily multivitamin (and various antioxidant supplements such as alpha lipoic acid) can also help maintain your body's energy release mechanisms at the cellular level.

Blood sugar is typically lowest first thing in the morning, so this is a lousy time to actively study. On the other hand, a good hearty breakfast, eaten at least an hour before plunging in to your studies, is a great idea and is a far better habit to develop for a lifetime than eating no breakfast and instead eating a huge meal right before bed.

Learning requires adequate *sleep*. Sure this is tough to manage at college – there are no parents to tell you to go to bed, lots of things to do, and of course you're in *class* during the day and then you study, so late night is when you have fun. Unfortunately, learning is clearly correlated with engagement, activity, and mental alertness, and all of these tend to shut down when you're tired. Furthermore, the formation of *long term memory of any kind* from a day's experiences has been shown in both animal and human studies to *depend* on the brain undergoing at least a few natural sleep cycles of deep sleep alternating with REM (Rapid Eye Movement) sleep, dreaming sleep. Rats taught a maze and then deprived of REM sleep cannot run the maze well the next day; rats that are taught the *same* maze but that get a good night's worth of rat sleep with plenty of rat dreaming can run the maze well the next day. People conked on the head who remain unconscious for hours and are thereby deprived of normal sleep often have permanent amnesia of the previous day – it never gets turned into long term memory.

This is hardly surprising. Pure common sense and experience tell you that your brain

[4]Wikipedia: http://www.wikipedia.org/wiki/glycemic index. t

won't work too well if it is hungry and tired. Common sense (and yes, experience) will rapidly convince you that learning generally works better if you're not stoned or drunk when you study. Learning works *much* better when you have *time* to learn and haven't put everything off to the last minute. In fact, all of Maslow's hierarchy of needs[5] are important parameters that contribute to the probability of success in learning. Yes, we all have heard stories of famous intellectuals who arose from poverty and social isolation and overcame it all to excel, but nobody tells the story of the *millions who do not*. First of all, these rare individuals probably succeeded because they were genetically gifted and ***really, really cared***; second, think about the *probability* of success, all other things being equal. It's hard to devote much time to learning (say) physics if you have to struggle just to stay fed and housed and clothed and safe from your also struggling and sometimes violent neighbors, have no mentor/instructor, and cannot easily afford even the paper and pen required to work problems, let alone a decent textbook[6].

There is one more set of very important variables that strongly affect our ability to learn, and they are in some ways the least well understood. These are variables that describe you as an *individual*, that describe your *particular* brain and how it works. Pretty much everybody will learn better if they are self-actualized and fully and actively engaged, if the material they are trying to learn is available in a form that makes sense and clearly communicates the implicit patterns that enable efficient information compression and storage, and above all if they *care* about what they are studying and learning, if it has *value* to them.

But everybody is not the same, and the *optimal* learning strategy for one person is not going to be what works well, or even at all, for another. This is one of the things that confounds "simple" empirical research that attempts to find benefit in one teaching/learning methodology over another. Some students *do* improve, even dramatically improve – when this or that teaching/learning methodology is introduced. In others there is no change. Still others actually do worse. In the end, the beneficial effect to a selected subgroup of the students may be lost in the statistical noise of the study and the fact that no attempt is made to identify commonalities among students that succeed or fail.

The point is that finding an optimal teaching and learning strategy is *technically* an *optimization problem on a high dimensional space*. We've discussed *some* of the important dimensions above, isolating a few that appear to have a monotonic effect on the desired outcome in at least some range (relying on common sense to cut off that range or suggest trade-offs – one cannot learn better by simply discussing one idea for weeks at the expense of participating in lecture or discussing many other ideas of equal and coordinated importance; sleeping for twenty hours a day leaves little time for experience to fix into long term memory with all of that sleep).

[5]Wikipedia: http://www.wikipedia.org/wiki/Maslow's_hierarchy_of_needs. In a nutshell, in order to become *self-actualized* and realize your full potential in activities such as learning you need to have your physiological needs met, you need to be safe, you need to be loved and secure in the world, you need to have good self-esteem and the esteem of others. Only then is it particularly *likely* that you can become self-actualized and become a great learner and problem solver, although of course there are always exceptions, learning heroes that emerge from adversity...

[6]This is why I have written this textbook and made it freely available to anyone who can manage to access it via the internet. I grew up in India back in the 1960's, surrounded by intelligent, hardworking, but *incredibly poor* people. Physics textbooks these days typically cost well over \$100 in hard cover form, and even *renting* online copies of textbooks for a *single semester* is a significant fraction of this, placing them absurdly out of the reach of literally millions of potential students around the world (and making their purchase a bit of a burden even to the comparatively wealthy students in developed countries) while supporting an entire ecology of ten-or-more percenters between the author(s) and the student(s), all of whom typically make *more money* than the author. Seriously. Is this sane? Is this how we make the world a better place for all to live in?

We've omitted one that is crucial, however. That is *your brain!*

1.3: Your Brain and Learning

Your brain is more than just a unique instrument. In some sense it is you. You could imagine having your brain removed from your body and being hooked up to machinary that provided it with sight, sound, and touch in such a way that "you" remain[7]. It is difficult to imagine that you still exist in any meaningful sense if your brain is taken out of your body and destroyed while your body is artificially kept alive.

Your brain, however, *is* an instrument. It has internal structure. It uses energy. It does "work". It is, in fact, a biological machine of sublime complexity and subtlety, one of the true wonders of the world! Note that this statement can be made quite independent of whether "you" are your brain per se or a spiritual being who happens to be using it (a debate that need not concern us at this time, however much fun it might be to get into it) – either way the brain itself is quite marvelous.

For all of that, few indeed are the people who bother to learn to actually *use* their brain effectively *as* an instrument. It just works, after all, whether or not we do this. Which is fine. If you want to get the most mileage out of it, however, it helps to read the manual.

So here's at least *one* user manual for your brain. It is by no means complete or authoritative, but it should be enough to get you started, to help you discover that you are actually a lot smarter than you think, once you realize that you can *change* the way you think and learn and experience life and gradually *improve* it.

In the spirit of the learning methodology that we eventually hope to adopt, let's simply itemize in no particular order the various features of the brain[8] that bear on the process of learning. Bear in mind that such a minimal presentation is more of a *metaphor* than anything else because simple (and extremely common) generalizations such as "creativity is a right-brain function" are not strictly true as the brain is far more complex than that.

- The brain is *bicameral*: it has two *cerebral hemispheres*[9] , right and left, with brain functions *asymmetrically* split up between them.
- The brain's hemispheres are connected by a networked membrane called the *corpus callosum* that is how the two halves talk to each other.
- The human brain consists of *layers* with a structure that recapitulates evolutionary phylogeny; that is, the core structures are found in very primitive animals and are common to nearly all vertebrate animals, with new layers (apparently) added by evolution on top of this core as the various phyla differentiated, fish, amphibian, reptile, mammal, primate, human. The outermost layer where most actual thinking occurs (in animals that think) is known as the *cerebral cortex*.

[7] Imagine very easily if you've ever seen *The Matrix* movie trilogy...
[8] Wikipedia: http://www.wikipedia.org/wiki/brain.
[9] Wikipedia: http://www.wikipedia.org/wiki/cerebral_hemisphere.

- The *cerebral cortex*[10] – especially the *outermost* outermost layer of *it* called the *neocortex* – is where "higher thought" activities associated with learning and problem solving take place, although the brain is a very complex instrument with functions spread out over many regions.

- An important brain model is a *neural network*[11] . Computer simulated neural networks provide us with insight into how the brain can remember past events and process new information.

- The fundamental operational units of the brain's information processing functionality are called *neurons*[12] . Neurons receive electrochemical signals from other neurons that are transmitted through long fibers called *axons*[13] *Neurotransmitters*[14] are the actual chemicals responsible for the triggered functioning of neurons and hence the neural network in the cortex that spans the halves of the brain.

- Parts of the cortex are devoted to the senses. These parts often contain a *map* of sorts of the world as seen by the associated sense mechanism. For example, there exists a topographic map in the brain that roughly corresponds to points in the retina, which in turn are stimulated by an image of the outside world that is projected onto the retina by your eye's lens in a way we will learn about later in this course! There is thus a *representation of your visual field* laid out inside your brain!

- Similar maps exist for the other senses, although sensations from the right side of your body are generally processed in a laterally inverted way by the *opposite* hemisphere of the brain. What your right eye sees, what your right hand touches, is ultimately transmitted to a sensory area in your left brain hemisphere and vice versa, and volitional muscle control flows from these brain halves the other way.

- Neurotransmitters require biological resources to produce and consume bioenergy (provided as glucose) in their operation. You can *exhaust* the resources, and *saturate* the receptors for the various neurotransmitters on the neurons by overstimulation.

- You can also block neurotransmitters by chemical means, put neurotransmitter analogues into your system, and alter the chemical trigger potentials of your neurons by taking various drugs, poisons, or hormones. The *biochemistry of your brain* is extremely important to its function, and (unfortunately) is not infrequently a bit "out of whack" for many individuals, resulting in e.g. attention deficit or mood disorders that can greatly affect one's ability to easily learn while leaving one otherwise highly functional.

- Intelligence[15] , learning ability, and problem solving capabilities are not fixed; they can vary (often improving) over your whole lifetime! Your brain is highly *plastic* and can sometimes even reprogram itself to full functionality when it is e.g. damaged by a stroke or accident. On the other hand neither is it infinitely plastic – any given brain has a

[10]Wikipedia: http://www.wikipedia.org/wiki/Cerebral_cortex.
[11]Wikipedia: http://www.wikipedia.org/wiki/Neural_network.
[12]Wikipedia: http://www.wikipedia.org/wiki/Neurons.
[13]Wikipedia: http://www.wikipedia.org/wiki/axon. .
[14]Wikipedia: http://www.wikipedia.org/wiki/neurotransmitters.
[15]Wikipedia: http://www.wikipedia.org/wiki/intelligence.

range of accessible capabilities and can be improved only to a certain point. However, for people of supposedly "normal" intelligence and above, it is by no means clear what that point is! Note well that *intelligence is an extremely controversial subject* and you should not take things like your own measured "IQ" too seriously.

- Intelligence is not even fixed within a population over time. A phenomenon known as "the Flynn effect"[16] (after its discoverer) suggests that IQ tests have increased almost six points a decade, on average, over a timescale of tens of years, with most of the increases coming from the lower half of the distribution of intelligence. This is an active area of research (as one might well imagine) and some of that research has demonstrated fairly conclusively that individual intelligences can be improved by five to ten points (a significant amount) by environmentally correlated factors such as nutrition, education, complexity of environment (empirically validating my earlier assertion that these factors *do* significantly affect the probability of long term success at learning).

- The best time for the brain to learn is right before sleep. The process of sleep appears to "fix" long term memories in the brain and things one studies right before going to bed are retained much better than things studied first thing in the morning. Note that this conflicts directly with the party/entertainment schedule of many students, who tend to study early in the evening and then amuse themselves until bedtime. It works much better the other way around, with strong limits on chemical "amusements".

- Sensory memory[17] corresponds to the roughly 0.5 second (for most people) that a sensory impression remains in the brain's "active sensory register", the sensory cortex. It can typically hold less than 12 "objects" that can be retrieved. It quickly decays and cannot be improved by rehearsal, although there is some evidence that its object capacity can be improved over a longer term by practice.

- Short term memory is where *some* of the information that comes into sensory memory is transferred. Just which information is transferred depends on where one's "attention" is, and the mechanics of the attention process are not well understood and are an area of active research. Attention acts like a filtering process, as there is a *wealth* of parallel information in our sensory memory at any given instant in time but the thread of our awareness and experience of time is serial. We tend to "pay attention" to one thing at a time, although some people can split their attention up to some extent (driving and talking, gardening and listening to a story or music) as long as the attention is split into different cognitive "channels" (that is, not reading and listening to a story or playing a waltz while listening to rock and roll) or can serialize their attention rapidly while retaining subtask memory not unlike the way single processor computers multitask (e.g. chess masters playing twenty games "at once").

  Short term memory lasts from a few seconds to as long as a minute without rehearsal, and for nearly all people it holds $4-5$ objects[18]. However, its capacity can be increased by a process called "chunking" that is basically the information compression mechanism

[16] Wikipedia: http://www.wikipedia.org/wiki/flynn_effect.

[17] Wikipedia: http://www.wikipedia.org/wiki/memory. Several items in a row are connected to this page.

[18] From this you can see why I used ten digits, gave you only a few seconds to look, and blocked rehearsal in our earlier exercise.

demonstrated in the earlier example with numbers – grouping of the data to be recalled into "objects" that permit a larger set to still fit in short term memory.

- Studies of chunking show that the ideal size for data chunking is three. That is, if you try to remember the string of letters:

  FBINSACIAIBMATTMSN

  with the usual three second look you'll almost certainly find it impossible. If, however, I insert the following spaces:

  FBI NSA CIA IBM ATT MSN

  It is suddenly much easier to get at least the first four. If I parenthesize:

  (FBI NSA CIA) (IBM ATT MSN)

  so that you can recognize the first three are all government agencies in the general category of "intelligence and law enforcement" and the last three are all market symbols for information technology mega-corporations, you can once again recall the information a day later with only the most cursory of rehearsals. You've taken eighteen "random" objects that were meaningless and could hence be recalled only through the most arduous of rehearsal processes, converted them to six "chunks" of three in a way that can be easily tagged by the brain's existing long term memory (note that you are *not learning* the string FBI, you are building an *association* to the already existing memory of what the string FBI *means*, which is *much easier* for the brain to do), and then chunked the chunks into *two* objects – well-known government security services and well-known megacorporation nicknames.

  Eighteen objects without meaning – difficult indeed! Those *same* eighteen objects *with* meaning – umm, looks pretty easy, doesn't it...

  Short term memory is still that – short term. It typically decays on a time scale that ranges from minutes for nearly everything to order of a day for a few things unless the information can be transferred to *long* term memory. Long term memory is the big payoff – *learning* is associated with formation of long term memory.

- Now we get to the really good stuff. Long term is memory that you form that lasts a long time in human terms. A "long time" can be days, weeks, months, years, or a lifetime. Long term memory is encoded *completely differently* from short term or sensory/immediate memory – it appears to be encoded *semantically*[19] , that is to say, *associatively* in terms of its *meaning*. There is considerable evidence for this, and it is one reason we focus so much on the importance of meaning in the previous sections.

  To miraculously transform things we try to remember from "difficult" to learn random factoids that have to be brute-force stuffed into disconnected semantic storage units created as it were one at a time for the task at hand into "easy" to learn factoids, all we have to do is *discover* meaning associations with things we already know, or *create* a strong memory of the global meaning or *conceptualization* of a subject that serves as an associative home for all those little factoids.

  A characteristic of this as a successful process is that when one works systematically to learn by means of the latter process, learning gets *easier* as time goes on. Every

[19] Wikipedia: http://www.wikipedia.org/wiki/semantics.

factoid you add to the semantic structure of the global conceptualization strengthens it, and makes it even easier to add new factoids. In fact, the mind's extraordinary rational capacity permits it to interpolate and extrapolate, to *fill in* parts of the structure on its own *without effort* and in many cases without even being exposed to the information that needs to be "learned"!

- One area where this extrapolation is particularly evident and powerful is in *mathematics*. Any time we can learn, or discover from experience a *formula* for some phenomenon, a mathematical *pattern*, we don't have to actually see something to be able to "remember" it. Once again, it is easy to find examples. If I give you data from sales figures over a year such as January = $1000, October = $10,000, December = $12,000, March=$3000, May = $5000, February = $2000, September = $9000, June = $6000, November = $11,000, July = $7000, August = $8000, April = $4000, at first glance they look quite difficult to remember. If you organize them temporally by month and look at them for a moment, you recognize that sales increased *linearly* by month, starting at $1000 in January, and suddenly you can reduce the whole series to a simple mental formula (straight line) and a couple pieces of initial data (slope and starting point). One amazing thing about this is that if I asked you to "remember" something that you *have not seen*, such as sales in February in the *next* year, you could make a very plausible guess that they will be $14,000[20]!

  Note that this isn't a memory, it is a guess. Guessing is what the mind is designed to do, as it is part of the process by which it "predicts the future" even in the most mundane of ways. When I put ten dollars in my pocket and reach in my pocket for it later, I'm basically guessing, on the basis of my memory and experience, that I'll find ten dollars there. Maybe my guess is wrong – my pocket could have been picked[21], maybe it fell out through a hole. My *concept* of object permanence plus my *memory* of an initial state permit me to make a *predictive guess* about the Universe!

  This is, in fact, physics! This is what physics is all about – coming up with a set of rules (like conservation of matter) that encode observations of object permanence, more rules (equations of motion) that dictate how objects move around, and allow me to conclude that "I put a ten dollar bill, at rest, into my pocket, and objects at rest remain at rest. The matter the bill is made of cannot be created or destroyed and is bound together in a way that is unlikely to come apart over a period of days. Therefore the ten dollar bill is still there!" Nearly anything that you do or that happens in your everyday life can be formulated as a predictive physics problem because "Physics makes the world go round" is *not just a metaphor!*

- The *hippocampus*[22] appears to be partly responsible for both forming spatial maps or visualizations of your environment and also for forming the *cognitive map* that organizes what you know and transforms short term memory into long term memory, and it appears to do its job (as noted above) *in your sleep*. Sleep deprivation *prevents the formation of long term memory*. Being rendered unconscious for a long period often produces *short*

[20]...or, um, $2000? Maybe sales follow an *annual pattern*, not an *unlimited growth* pattern. Gee, there *could* be more than one model *even when we don't immediately see it or think of it!* Which is why science is *empirical* – observing $2000 in the next year would force us to reconsider the original model, wouldn't it?

[21]With three sons constantly looking for funds to attend movies and the like, it isn't as unlikely as you might think!

[22]Wikipedia: http://www.wikipedia.org/wiki/hippocampus.

term amnesia as the brain loses short term memory before it gets put into long term memory. The hippocampus shows evidence of plasticity – taxi drivers who have to learn to navigate large cities actually have larger than normal hippocampi, with a size proportional to the length of time they've been driving. This suggests (once again) that it is possible to *deliberately increase the capacity* of your *own* hippocampus through the exercise of its functions, and consequently *increase your ability to store and retrieve information*, which is an important component (although not the only component) of intelligence!

- Memory is improved by *increasing the supply of oxygen to the brain*, which is best accomplished by *exercise*. Unsurprisingly. Indeed, as noted above, having good general health, good nutrition, good oxygenation and perfusion – having all the biomechanism in tip-top running order – is perfectly reasonably linked to being able to perform at your best in anything, mental activity included.

- Finally, the *amygdala*[23] is a brain organ in our *limbic system* (part of our "old", reptile brain). The amygdala is an important part of our *emotional* system. It is associated with primitive survival responses, with sexual response, with addiction, and appears to play a *key role* in modulating (filtering) the process of turning short term memory into long term memory. Basically, any short term memory associated with a powerful emotion is much more likely to make it into long term memory.

  There are clear evolutionary advantages to this. If you narrowly escape being killed by a saber-toothed tiger at a particular pool in the forest, and then forget that this happened by the next day and return again to drink there, chances are decent that the saber-tooth is still there and you'll get eaten. On the other hand, if you come upon a particular fruit tree in that same forest and get a free meal of high quality food and forget about the tree a day later, you might starve.

  We see that both negative and positive emotional experiences are strongly correlated with learning! *Powerful* experiences, especially, are correlated with learning. This translates into learning strategies in two ways, one for the instructor and one for the student. For the instructor, there are two general strategies open to helping students learn. One is to create an atmosphere of *fear, hatred, disgust, anger* – powerful negative emotions. The other is to create an atmosphere of *love, security, humor, joy* – powerful positive emotions. In between there is a great wasteland of bo-ring, bo-ring, bo-ring where students plod along, struggling to form memories because there is nothing "exciting" about the course in either a positive or negative way and so their amygdala degrades the memory formation process in favor of other more "interesting" experiences.

  Now, in my opinion, negative experiences in the classroom do indeed promote the formation of long term memories, but they aren't the memories the instructor intended. The student is likely to remember, and loathe, the instructor for the rest of their life but is *not* more likely to remember the material except sporadically in association with particularly traumatic episodes. They may well be *less* likely, as we naturally avoid negative experiences and will study less and work less hard on things we can't stand doing or that make us feel bad about ourselves.

[23] Wikipedia: http://www.wikipedia.org/wiki/amygdala.

For the instructor, then, positive is the way to go. Creating a warm, nurturing classroom environment, ensuring that the students know that you *care* about their learning and about them as individuals helps to promote learning. Making your lectures and teaching processes *fun* – and *funny* – helps as well[24]. Many successful lecturers make a powerful *positive* impression on the students, creating an atmosphere of amazement or surprise. A classroom experience should really be a *joy* in order to optimize learning in so many ways.

For the student, be aware that *your attitude matters!* As noted in previous sections, *caring* is an essential component of successful learning because you have to attach *value* to the process in order to get your amygdala to do its job. However, you can do *much more*. You can see how *many* aspects of learning can be enhanced through the simple expedient of making it a positive experience! *Working in groups*, working with a team of peers, is *fun*, and you learn more when you're having fun (or quavering in abject fear, or in an interesting mix of the two) than when you are alone, disconnected and bored. Attending an *interesting* lecture is fun, and you'll retain more, on average, from an interesting lecture than you will from a boring one. *Participation in general* is fun, especially if you are "rewarded" in some way that makes a moment or two special to you, and you'll remember more of what goes on.

Chicken or egg? We see a fellow student who is relaxed and appears to be having fun because they are doing really well in the course where we are constantly stressed out and struggling, and we write their happiness off as being due to their success and our misery as being caused by our failure. It is possible, however, that we have this backwards! Perhaps they are doing really well in the course ***because they are relaxed and having fun***, perhaps we are doing not so well ***because for us, every minute in the classroom is a torture!***

In any event, you've probably tried misery in the classroom in at least one class already. How'd that work out for you? Perhaps it is worth trying joy, instead!

I'm hoping that all of these little factoids (presented in a way that ought to help you to build at least the beginnings of a working conceptual model of your own brain) are clearly communicating to you the subtext message that *all of this is at least partially under your control!* Even if your instructor is scary or boring, the material at first glance seems dry and meaningless, and so on – all the negative-neutral things that make learning difficult, *you* can decide to make it fun and exciting, *you* can ferret out the meaning, *you* can adopt study strategies that focus on the formation of cognitive maps and organizing structures *first* and *then* on applications, rehearsal, factoids, and so on, *you* can learn to study right before bed, get enough sleep, become aware of your brain's learning biorhythms.

Finally, you can learn to *increase your functional learning capabilities* by a *significant* amount. Solving puzzles, playing mental games, doing crossword puzzles or sudoku, working homework problems, writing papers, arguing and discussing, just plain *thinking* about difficult subjects and problems even when you don't *have* to all increase your active intelligence in

[24]...as long as the standup comedy routine doesn't compete with or interfere with the material that is the *object* of the learning process, of course. Again, we don't want our students to be fondly remembering Professor Bonkers for his many wonderful one-liners in class while they can no longer remember what *subject* he was actually teaching them after ten years.

initially small but cumulative ways. You too can increase the size of your hippocampus by navigating a new subject instead of a city, you too can learn to engage your amygdala by *choosing* in a self-actualized way what you value and learning to discipline your emotions accordingly, you too can create more conceptual maps within your brain that can be shared as components across the various things you wish to learn.

The more you know about anything, the easier it is to learn everything – and vice versa! This is a tenet of *epistemology*[25] *!* "Knowledge" is a network of interconnected beliefs selected for their mutual consistency and the degree of their congruence with the world of our experience. Of course, it is *also* the reason a liberal arts education is so valuable! If your knowledge is too narrow, it can easily be inconsistent with the world in places you are not familiar with. If it is too shallow the connections between topics that at first glance are quite different are not apparent. Only when it is broad and at least somewhat deep does learning itself become easier and easier as we ***get the big picture*** of what the state of the world itself is "right now" and how it operates, giving us at least a *chance* of choosing between the many futures that fork out of the now on the basis of our decisions and actions[26].

From this we can conclude that in order for you to learn physics efficiently, you should ***use your whole brain*** and ***exercise it often***. Don't think that you "just" need math and not spatial relations, visualization, verbal skills, a knowledge of history, a memory of performing experiments with your hands or mind or both – you need it all! Remember, just as is the case with physical exercise (which you should get plenty of), *mental* exercise gradually makes you mentally stronger, so that you can eventually do easily things that at first appear insurmountably difficult. You can learn to learn *three to ten times as fast* as you did in high school, to have more fun while doing it, and to gain tremendous reasoning capabilities along the way just by *trying* to learn to learn more efficiently instead of continuing to use learning strategies that worked (possibly indifferently) back in elementary and high school.

The next section, at long last, will make a very specific set of suggestions for *one* very good way to study physics (or nearly anything else) in a way that maximally takes advantage of your own volitional biology to make learning as efficient and pleasant as it is possible to be.

1.4: How to Do Your Homework Effectively

By now in your academic career (and given the information above) it should be very apparent just where homework exists in the grand scheme of (learning) things. Ideally, you attend a class where a warm and attentive professor clearly explains some abstruse concept and a whole raft of facts in some moderately interactive way that encourages engagement and "being earnest". Alas, there are *too many* facts to fit in short term/immediate memory and *too little time* to move most of them through into long term/working memory before finishing with one and moving

[25] Wikipedia: http://www.wikipedia.org/wiki/Epistemology. The study of knowledge itself – how we know what we know, just what "meaning" or "conceptual understanding" are. Ultimately, epistemology attempts to define *what it is best to believe* given our experience, the evidence obtained from the world itself.

[26] Sure, OK, maybe I'm getting carried away, as a lot of this involves *politics* and *ethics*, subjects not traditionally associated with physics and science. Or, maybe not! In my opinion, at least, *they should be*, as empiricism and common sense are just as important in sociopolitical ethical choices as they are in choosing to drive on the correct side of the road to avoid highly inelastic collisions with oncoming traffic.

on to the next one. The material may appear to be boring and random so that it is difficult to pay full attention to the *patterns* being communicated and remain emotionally enthusiastic all the while to help the process along. As a consequence, by the end of lecture you've already *forgotten* many if not most of the facts, but if you were paying attention, asked questions as needed, and really cared about learning the material you *would* remember a handful of the most important ones, the ones that made your brief understanding of the material hang (for a brief shining moment) together.

This conceptual overview, however initially tenuous, is the skeleton you will eventually clothe with facts and experiences that transform it into an entire system of associative memory and reasoning where you can work intellectually at a high level with little effort and usually with a great deal of pleasure associated with the very act of thinking. But you aren't there yet.

You now know that you are not terribly likely to retain a lot of what you are shown in lecture without engagement. In order to actually learn it, you must *stop* being a passive recipient of facts. You must *actively* develop your understanding, using methods such as: *preparing* for lecture (a little goes a long way towards *keeping* lecture from being overwhelming); *asking questions* during and after lecture; *discussing* the material and kicking it around with other students or mentors; *using* the material in some way as soon as possible to reinforce the concepts and reveal places where your understanding needs to be shored up by asking more questions; *teaching* the material to peers as you come to understand it; and:

Work towards *mastery*, not just *completion!*

Let's examine this latter point, especially, in some detail.

To help facilitate the learning process *begun* with lecture, your professor/teacher almost certainly gave you a ***homework assignment***. Amazingly enough, its purpose is *not* to torment you *or* to be used as a major basis of your grade (although it may well do both in the structured environment of a University class). Its purpose to ***give you some concrete stuff to do while thinking about the material to be learned, discussing the material to be learned, and using the material to be learned*** to accomplish specific goals of graduated difficulty, ideally while teaching some of what you figure out to others who are sharing this whole experience while being taught by them in turn.

The assignment – if your professor/teacher/mentor did their job well and gave you a *good* one, a *thoughtful* one, one that is *as difficult or more difficult than any planned quiz or exam on the material* – is *much more important* than lecture, as learning is something *you* do as a result of your actions and focused attention, not something that happens "automatically" because of something your professor says or does. Doing the assignment *properly* should – in my opinion – more or less guarantee that a student do *approximately as well on tests and quizzes as they did on the assignments!*

This last point is especially important. One of the most common mistakes a student makes – one that was very probably turned into inviolable habit way, way back in secondary school, often by the end of first grade – is to ***put the least possible amount of effort into your homework!*** At first, the motivation was obvious. You didn't get to go out and play until your homework was done. Or, you got into trouble if you didn't do it. Or, you were really tired when you got home, more than a bit bored with school in general, and didn't feel like doing still more work at all. Or, you *couldn't do* the homework well when you got home, and it was less ego-dystonic for you to not do it at all or to deliberately do it minimally to give yourself an "excuse" for poor performance.

No matter what the reason, you very likely built up a *habit* of putting in *just enough work to – barely – complete the homework* by going back and forth in the book and perhaps using your notes and examples *just enough* get you through the assignment, possibly copying or minimally adapting from the examples (or worse, from other, possibly better, students).

I'm going to ask you here, and I want you to be cruelly, bluntly honest in your self-appraised answer: ***In what Universe could you reasonably expect this to work?*** The whole *point* of homework is to give you guided, scaled tasks that, when done to the point of *mastery* – the ability to do the homework perfectly ***without looking at the book, notes, or anything at all, a day or more after completing the assignment*** – enables you to really, truly, conceptually learn the material you are nominally studying as you go along, not just doing any required in-class or home work *to completion* while telling yourself that you'll figure out all of the stuff you still don't remember and/or understand *later*, just in time for the big test(s) on which most of your grade depends.

Here's my own experience, speaking *as a student* and not as a professor. Later never comes. When your professor puts something up on the board or screen that you don't understand and you don't stop them and ask a question to help clarify it, telling yourself that you'll go over the notes and book *later* and figure it out – later never comes. When you get your homework done only because a friend or the teaching assistant or the help room mentor or the professor him/her/itself showed you how to work one or more of the problems and you hand it in telling yourself that you'll do it again *later* – later never comes. When you get the homework back, and discover you got a third of the problems wrong (and you still don't really see *why* you got them wrong or *how* to do them correctly) but instead of redoing them until you get them right, perfectly, without looking (bugging the grader, the professor, a good friend who got them right for help as needed) *right now*, you tell yourself that you'll figure it out and work them correctly *later* – later never comes.

Or, rather, it does come. It comes at the very end of the class, 2 to 3 days before the final exam, when you realize with a cold feeling in the pit of your stomach that "I have not done that which I ought to have done, and am sore afraid" and now there is no longer *time* to do them given the long, long list you accumulated over the entire course. So you spend time ineffectually trying to memorize a bunch of stuff because there is no way you will actually *learn* it in 3 days, especially with other classes with *other* work you blew off until – later. Not a good look.

This is so easily avoided, and avoiding it can utterly transform your academic experience. Stop viewing homework as something annoying you have to do instead of playing frisbee with friends, rocking World of Warcraft or Diablo IV, partying, or just sitting around and *not* doing homework while also not doing anything else of any particular interest. View it as being *even more critical to your grade and actual learning than going to class*. Because it is! Over the decades I have taught physics, I have known many students who were highly indifferent in their class attendance but who worked very hard on their homework and keeping up with the material who succeeded, and far *more* students who have shown up in class fairly regularly, but who didn't do their homework or did it badly, just enough to get by, who got well-(un)earned terrible grades.

If you fail to do the assignments *with your entire spirit engaged*, working until you can do the assignments without any external support (such as looking in the textbook, cheating, getting help from friends, etc) how can you be surprised if you can't do the quizzes and exams where the problems are (no suprise) *just like* the homework problems you couldn't do if they *were* the exam questions!

In other words, to learn you must *do your homework*, where by "do" I really mean "doing in such a way that you master it" and *not just finishing it, handing it in, and forgetting it!*

Well now, you might well ask. How do I do *that*, especially when I've never really done it before and don't really know how? How do I avoid the frustration and anger that can easily arise when I really *do* try to master it and it *doesn't work?*.

The first piece of advice I have for you here is to work at least partly in a *team* setting. I'm a huge fan of *team-based learning*. Being a member of a team or homemade study group does not absolve you from the duty of mastering the homework – properly done, it enables it. If your teacher uses team-based learning that mixes lecture and in-class structured mentored problem solving practice as "pre"-homework work, so much the better, but whether or not they do doesn't affect *your* choices of how to best study.

Beyond that, let's concentrate on "best practices" for individual study, as at least part of mastery is working alone to master it *alone*. There are two general steps that need to be *iterated* to finish learning anything at all. I'm not going to lie to you: They are a lot of work. In fact, they are likely to be *more* work than (passively) attending lecture, and are (as I already noted above) *more important* than attending lecture. You can learn the material with these steps without *ever* attending lecture, as long as you have access to what you need to learn in some media or human form. You in all probability will *never* learn it, lecture or not, without making a few passes through these steps. They are:

a) Review the whole (typically textbooks and/or notes)

b) Work on the parts (do homework, use it for something)

(iterate until you thoroughly understand whatever it is you are trying to learn).

Let's examine these steps.

The first is pretty obvious. You didn't "get it" from one lecture. There was too much material. If you were *lucky* and well prepared and blessed with a good instructor, perhaps you grasped *some* of it for a *moment* (and if your instructor was poor or you were particularly poorly prepared you may not have managed even that) but what you did momentarily understand is fading, flitting further and further away with every moment that passes. You need to review the entire topic, as a whole, as well as all its parts. A set of good summary notes might contain all the relative factoids, but there are *relations* between those factoids – a temporal sequencing, mathematical derivations connecting them to other things you know, a topical association with other things that you know. They tell a *story*, or part of a story, and you need to know that story in *broad* terms, not try to memorize it word for word. On the good side, work or not, stories are *fun* where memorizing piles of facts is *not!*

Reviewing the material should be done in layers, skimming the textbook and your notes, creating a *new* set of notes out of the text in combination with your lecture notes, maybe reading in more detail to understand some particular point that puzzles you, reworking a few of the examples presented. Lots of increasingly deep passes through it (starting with the merest skim-reading or reading a summary of the whole thing) are *much* better than trying to work through the whole text one line at a time and not moving on until you understand it. Many things you might want to understand will only come clear from things you are exposed to *later*, as it is not the case that all knowledge is ordinal, hierarchical, and derivative.

You especially do *not* have to work on *memorizing* the content. In fact, it is *not* desireable to try to memorize content at this point – you want the big picture *first* so that facts have a place to live in your brain. If you build them a house, they'll move right in without a fuss, where if you try to grasp them one at a time with no place to put them, they'll (metaphorically) slip away again as fast as you try to take up the next one. Let's understand this a bit.

As we've seen, your brain is fabulously efficient at storing information in a *compressed associative* form. It also tends to remember things that are *important* – whatever that means – and forget things that aren't important to make room for more important stuff, as your brain structures work together in understandable ways on the process. Building the cognitive map, the "house", is what it's all about. But as it turns out, building this house *takes time*.

This is the goal of your iterated review process. At first you are memorizing things the hard way, trying to connect what you learn to very simple hierarchical concepts such as this step comes before that step. As you do this over and over again, though, you find that absorbing new information takes you less and less time, and you remember it much more easily and for a longer time without additional rehearsal. Sometimes your brain even *outruns* the learning process and "discovers" a missing part of the structure before you even read about it! By reviewing the whole, well-organized structure over and over again, you gradually build a greatly compressed representation of it in your brain and tremendously reduce the amount of work required to flesh out that structure with increasing levels of detail *and remember them and be able to work with them* for a long, long time.

Now let's understand the second part of doing homework – working problems. As you can probably guess on your own at this point, there are good ways and bad ways to do homework problems. The worst way to do homework (aside from not doing it at all, which is *far too common* a practice and a *bad idea* if you have any intention of learning the material) is to do it all in one sitting, right before it is due, and to never again look at it.

Doing your homework in a single sitting, working on it just one time *fails to repeat and rehearse the material* (essential for turning short term memory into long term in nearly all cases). It *exhausts the neurons in your brain* (quite literally – there is metabolic energy consumed in thinking) as one often ends up working on a problem far too long in one sitting just to get done. It *fails to incrementally build up* in your brain's long term memory the *structures* upon which the more complex solutions are based, so you have to constantly go back to the book to get them into short term memory long enough to get through a problem. Even this simple bit of repetition does *initiate* a learning process. Unfortunately, by not repeating them after this one sitting they soon fade, often without a discernable trace in long term memory.

Just as was the case in our experiment with memorizing the number above, the problems almost invariably are *not* going to be a matter of random noise. They have certain key facts and ideas that are the basis of their solution, and those ideas are used over and over again. There is plenty of pattern and meaning there for your brain to exploit in information compression, and it may well be *very cool stuff to know* and hence *important* to you once learned, but it takes time and repetition and a certain amount of meditation for the "gestalt" of it to spring into your awareness and burn itself into your conceptual memory as "high order understanding".

You have to *give* it this time, and perform the repetitions, while maintaining an optimistic, philosophical attitude towards the process. You have to do your best to have *fun* with it. You don't get strong by lifting light weights a single time. You get strong lifting weights repeatedly, starting with light weights to be sure, but then working up to the *heaviest weights you can manage*. When you *do* build up to where you're lifting hundreds of pounds, the fifty pounds you started with seems light as a feather to you.

As with the body, so with the brain. Repeat broad strokes for the big picture with increasingly deep and "heavy" excursions into the material to explore it in detail as the overall picture emerges. Intersperse this with sessions where you *work on problems* and try to *use* the material you've figured out so far. Be sure to *discuss* it and *teach it to others* as you go as much as possible, as articulating what you've figured out to others both uses a different part of your brain than taking it in (and hence solidifies the memory) and it helps you articulate the ideas to *yourself!* This process will help you learn more, better, faster than you ever have before, and to have fun doing it!

Your brain is more complicated than you think. You are very likely used to *working hard* to try to *make* it figure things out, but you've probably observed that this doesn't work very well. A lot of times you simply *cannot* "figure things out" because your brain doesn't yet know the key things required to do this, or doesn't "see" how those parts you do know fit together. Learning and discovery is not, alas, "intentional" – it is more like trying to get a bird to light on your hand that flits away the moment you try to grasp it.

People who do really hard crossword puzzles (one form of great brain exercise) have learned the following. After making a pass through the puzzle and filling in all the words they can "get", and maybe making a couple of extra passes through thinking hard about ones they can't get right away, looking for patterns, trying partial guesses, they arrive at an impasse. If they continue working hard on it, they are unlikely to make further progress, no matter how long they stare at it.

On the other hand, if they *put the puzzle down* and *do something else for a while* – especially if the something else is go to bed and sleep – when they come back to the puzzle they often can *immediately see* a dozen or more words that the day before were absolutely invisible to them. Sometimes one of the *long theme answers* (perhaps 25 characters long) where they have no more than *two letters* just "gives up" – they can simply "see" what the answer must be.

Where do these answers come from? The person has not "figured them out", they have "recognized" them. They come all at once, and they don't come about as the result of a logical sequential process.

Often they come from the person's *right brain*[27]. The left brain tries to use logic and simple

[27]Note that this description is at least partly metaphor, for while there is some hemispherical specialization of some of these functions, it isn't always sharp. I'm retaining them here (oh you brain specialists who might be reading this) because they are a *valuable* metaphor.

memory when it works on crosswork puzzles. This is usually good for some words, but for many of the words there are *many possible answers* and without any insight one can't even recall *one* of the possibilities. The clues don't suffice to connect you up to a word. Even as letters get filled in this continues to be the case, not because you don't *know* the word (although in really hard puzzles this can sometimes be the case) but because you don't know how to *recognize* the word "all at once" from a cleverly nonlinear clue and a few letters in this context.

The right brain is (to some extent) responsible for *insight* and *non-linear thinking*. It sees *patterns*, and *wholes*, not sequential relations between the parts. It isn't intentional – we can't "make" our right brains figure something out, it is often the other way around! Working hard on a problem, then "sleeping on it" (to get that all important hippocampal involvement going) is actually a *great* way to develop "insight" that lets you solve it *without really working terribly hard* after a few tries. It also utilizes more of your brain – left and right brain, sequential reasoning and insight, and if you articulate it, or use it, or make something with your hands, then it exercises these parts of your brain as well, strengthening the memory and your understanding still more.

Make no mistake about it: the learning that is associated with this iterative process, and the problem solving power of the method, is *much greater* than just working on a problem linearly the night before it is due until you hack your way through it using information assembled a part at a time from the book.

The following "Method of Three Passes" is a *specific* strategy that implements many of the tricks discussed above. It is known to be effective for learning by means of doing homework (or in a generalized way, learning anything at all). It is ideal for "problem oriented homework", and will pay off big in learning dividends should you adopt it, especially when supported by a *group/team oriented pass mixed in* that provides *strong tutorial support* and *many opportunities for peer discussion and teaching*.

The Method of Three Passes

The following is presented in a way slightly specialized to learning the physics in this textbook, but is (hopefully obviously) easily generalized to work in other classes, perhaps with slightly modified patterns of study.

Pass 1 Three or more nights before recitation (or when the homework is due), make a *fast* pass through all problems. Plan to spend 1-1.5 hours on this pass. With roughly 10-12 problems, this gives you around 6-8 minutes per problem. Spend *no more* than this much time *per problem* and if you can solve them in this much time fine, otherwise move on to the next. Try to do this the last thing before bed at night (seriously) and *then go to sleep*.

Pass 2 After at least one night's sleep, make a *medium speed* pass through all problems. Plan to spend 1-1.5 hours on this pass as well. Some of the problems will already be solved from the first pass or nearly so. *Quickly* review their solution and then move on to concentrate on the still unsolved problems. If you solved 1/4 to 1/3 of the problems in the first pass, you should be able to spend 10 minutes or so per problem in the second pass. Again, do this right before bed if possible and then go immediately to sleep.

Pass 3 After at least one night's sleep, make a *final* pass through all the problems. Begin as before by quickly reviewing all the problems you solved in the previous two passes. Then spend fifteen minutes or more (as needed) to solve the remaining unsolved problems. Leave any "impossible" problems for recitation – there should be no more than three from any given assignment, as a general rule. Go immediately to bed.

This is an *extremely powerful* prescription for deeply learning nearly *anything*. Here is the motivation. Memory is formed by repetition, and this obviously contains a lot of that. Permanent (long term) memory is actually formed in your sleep, and studies have shown that whatever you study right before sleep is most likely to be retained. Physics is actually a "whole brain" subject – it requires a synthesis of both right brain visualization and conceptualization and left brain verbal/analytical processing – both geometry and algebra, if you like, and you'll often find that problems that stumped you the night before just solve themselves "like magic" on the second or third pass if you work hard on them for a short, intense, session and then sleep on it. This is your right (nonverbal) brain participating as it develops intuition to guide your left brain algebraic engine.

Other suggestions to improve learning include working in a study group or team for that third pass (the first one or two are best done alone to "prepare" for the third pass). Teaching is one of the best ways to learn, and by working in a group you'll have opportunities to both teach and learn more deeply than you would otherwise as you have to articulate your solutions.

Make the learning *fun* – the *right* brain is the key to forming long term memory and it is the seat of your *emotions*. If you are happy studying and make it a positive experience, you will increase retention, it is that simple. Again, working at least one pass with a group or team, making it a social experience, sharing the struggle, sharing the satisfaction one feels when you finally "get it", even just sharing the misery you feel when you *don't* eventually get it and have to seek help from a mentor, is *obviously* a way to make it a *lot* more fun. Order pizza, play music, make it a "physics homework party night".

Try to use your *whole brain* on the problems – draw lots of pictures and figures (right brain) to go with the algebra (left brain). Listen to quiet music (right brain) while thinking through the sequences of events in the problem (left brain). Build little "demos" of problems where possible – even using your hands in this way helps strengthen memory.

Avoid memorization. You will learn physics far better if you learn to *solve* problems and *understand* the concepts rather than attempt to *memorize* the umpty-zillion formulas, factoids, and specific problems or examples covered at one time or another in the class. That isn't to say that you shouldn't learn the important formulas, Laws of Nature, and all of that – it's just that the learning should generally *not* consist of putting them on a big sheet of paper all jumbled together and then trying to memorize them as abstract collections of symbols out of context.

Be sure to review the problems *one last time* when you get your graded homework back. Learn from your mistakes or you will, as they say, be doomed to repeat them.

If you follow this prescription, you will have seen *every assigned homework problem* a minimum of five times before you take an actual exam on that material – three original passes doing the work, a final write up pass without looking to guarantee mastery and make it "pretty" to hand in, and a review pass when you get it back to learn from your mistakes. At least two of these should occur after you have solved *all* of the problems correctly the first time. If you follow this pattern (or anything *close* to this pattern) then when the time comes to study for exams, it should really be (for once) a *review process*, not a *cram*. Every problem will be like an old friend, and a very brief review will form a *sixth or seventh* pass through the assigned homework.

With this methodology (enhanced as required by the physics resource rooms, tutors, and help from your instructors) there is no reason for you do poorly in the course and every reason to expect that you will do well, perhaps very well indeed! And if you distribute this work intelligently, you'll can very likely manage this spending only the 3 to 7 hours per week on homework that is expected of you in any college course of this level of difficulty!

This ends our discussion of course preliminaries (for nearly *any* serious course you might take, not just physics courses) and it is time to get on with the actual material for *this* course.

1.5: Mathematics

Physics, as was noted in the preface, requires a solid knowledge of all mathematics through calculus. That's right, the whole nine yards: number theory, algebra, geometry, trigonometry, vectors, differential calculus, integral calculus, even a smattering of differential equations. You may well be reading this book intending to use it to learn physics (either on your own – good for you! – or in an actual class) *never* having taken a class in calculus, or perhaps you took a class in calculus but barely squeaked by with a C- or D+.

I ***strongly advise against attempting this!*** At Duke University (where I teach this course) calculus is a ***strict prerequisite*** for taking calculus based physics. This is for two reasons: First, a calculus class may well be the only place where you are exposed to things like series, summation symbols, vectors (including vector products) that are often not taught in high school algebra and trigonometry classes. Second, Newton *invented* calculus just so that he could *invent* a consistent and successful theory of physics.

Learning physics without calculus (as it is taught in most high school physics classes and sadly, in some University-level physics classes) is in my opinion a nearly complete waste of time. It becomes an exercise in the memorization of formulas because one can literally not understand where anything comes from or how it all fits together without calculus. It tends to concentrate on constant force/constant acceleration problems simply because they are pretty much the only ones one *can* solve without calculus, to such a fault that I usually have to help students *unlearn* the algebraic solutions they memorized in high school if this is their only exposure to physics before we can move forward and learn physics correctly.

A final problem is that physics based only on algebraic solutions to constant acceleration kinematics is *boring*, and students understandably come out of such a course bored and frustrated with physics through no fault of the discipline (which is anything but boring). Learning physics is hard work, without doubt. It can be frustrating if it is poorly taught, taught as an exercise in memorization and graded primarily on how well you can do the ***simple arithmetic*** of substituting this or that set of numbers into a memorized formula in a single step. Substantial research in the teaching and learning of physics have demonstrated that there is a huge gap between achieving *conceptual understanding of even the most elementary physics* and this strictly algebra+arithmetic approach to studying physics.

Of course, even students who *have* taken calculus successfully can have a bit of a gap there as well. Many calculus classes – perhaps understandably, perhaps not, I don't want to judge too harshly – concentrate for better or for worse on *skills*, on the *algebraic manipulations* of calculus. Those courses are easy to identify – they had a student doing regular homework assignments consisting of page after page of taking derivatives of that, doing integrals of another thing, without *one single exercise having the slightest bit of meaning!* Again it is all too common for students to have treated the course as an exercise in memorization more than an invitation of mastery even of the limited tools required to do most of the problems, an understandable student response when presented with what appears to be an overwhelming mountain of *meaningless* symbols instead of a much smaller mountain of symbols that actually mean something and have some use when you are done.

This creates two distinct problems when students start to learn physics, with calculus. First, well over 95% of the calculus problems they were drilled on turn out to be useless in any scientific discipline and are 'valuable' only as a hobby for people who like to solve for analytical derivatives or integrals of functions that don't represent anything whatsoever in the real world. Second, because of the tremendous dilution of efforts wasted on these obscure and – to anyone but a future math major – useless problems, they end up *inadequately* skilled in the 5% of the introductory calculus problems they might have studied that *are really, really valuable* in real world problems.

Let me be very specific. One can succeed in – nay, thrive! – in introductory physics if you have *truly mastered* only five basic integration/differentiation rules, plus the product rule for differentiation (and the corresponding integration by parts rule for integration), plus the chain rule for differentiation (and the corresponding u-substitution rule for integration). If I were to add anything to that list it would be the hyperbolic differentials (and integrals). Let's list them (as indefinite integrals, although some of them, such as the natural logs, are more useful as definite integrals when it comes to handling the physical dimensions of the arguments):

- $\dfrac{du^n}{du} = nu^{n-1}$ and $\displaystyle\int u^n du = \frac{1}{n+1}u^{n+1} + C$
- $\dfrac{d\ln|u|}{du} = \dfrac{1}{u}$ and $\displaystyle\int \frac{du}{u} = \ln|u| + C$
- $\dfrac{de^u}{du} = e^u$ and $\displaystyle\int e^u du = e^u + C$
- $\dfrac{d\sin u}{du} = \cos u$ and $\displaystyle\int \cos u\; du = \sin u + C$
- $\dfrac{d\cos u}{du} = -\sin u$ and $\displaystyle\int \sin u\; du = -\cos u + C$

plus the two "extra" formulas (that can be easily enough derived from the exponential rule above, as can the trig integrals for that matter):

- $\dfrac{d\sinh u}{du} = \cosh u$ and $\displaystyle\int \cosh u\; du = \sinh u + C$
- $\dfrac{d\cosh u}{du} = -\sinh u$ and $\displaystyle\int \sinh u\; du = -\cosh u + C$

where $\sinh u = \frac{1}{2}e^u - e^{-u}$ and $\cosh u = \frac{1}{2}e^u + e^{-u}$. Tangent and cotangent are just a matter of using the product rule, as are the hyperbolic equivalents. Trig substitution derivatives and integrals are similarly just a matter using the chain rule or u-substitution (but with some clever pictures that allow one to visualize them as e.g. triangles). It also helps to at least know about the Taylor series expansion of smooth functions and the slightly more specific binomial expansion, and the idea of convergence. The actual *use* of all of this, extended right into the evaluation of multidimensional integrals and derivatives as needed, is taught in a self contained way in a typical course, as physics instructors have long since learned not to rely on the calculus supposedly learned in calculus classes by incoming students.

To put it another way, the fundamental calculus formulas needed to *completely master* a one year (two semester) introductory physics sequence – mechanics, electricity, magnetism, optics, and diverse applications of all of the above all ***entirely based on calculus*** – can easily fit on one single page.

That isn't to say that memorizing that page is sufficient preparation in math, though. Physics builds on skills in algebra, geometry, trigonometry, the idea of vector spaces and vector decomposition, some familiarity with at least three separable coordinate frames (much of which, again, is taught or retaught as need be in the physics courses themselves), series, complex numbers – it really *uses* mathematics throughout, where every single mathematical tool and idea used has *meaning* and is *not* an empty exercise with meaningless symbols.

The skill that is in some sense the *least* important is – perhaps surprisingly, given the public perception of the discipline – ***arithmetic***. That isn't to say that physicists don't care about numbers or that you can get by in a physics class when you are unable to add, subtract, multiply or divide numbers with nothing but a pen and piece of paper, it's just recognition of the fact that once one understands what is going on and can solve problems correctly algebraically, plugging in the numbers is a *trivial final step*, one that can easily be done with a calculator or computer if need be, where even an arithmetical savant, somebody capable of multiplying 8 digit numbers instantly in their head, is helpless in physics if they cannot actually perform the conceptual reasoning, visualization, algebra, and dimensional analysis required to take a word problem, convert it into a picture with coordinates attached, decorate it with given forces or interactions, express the whole thing in the algebraic forms associated with the relevant physical laws, and then use *all* of the mathematical and algebraic skills one needs to obtain an algebraic solution that can be checked with dimensional analysis for consistency and rechecked with some simple does-it-make-sense rules. Sure, once that formula is obtained, they might be better and faster than a computer, but given the formula, a sixth grader with the same numerical data and a calculator can get just as accurate a solution in only a little more time, provided only that they know what those trig functions and so on on their calculator are good for.

In any event, if you are preparing to study calculus-based physics (from this book or any other), here is a list of a few of the kinds of things you'll have to be able to do during the next two semesters of physics. Don't worry just yet about what they *mean* – that is part of what you will learn along the way. The question is, can you (perhaps with a short review of things you've learned and knew at one time but have not forgotten) evaluate these mathematical expressions or solve for the algebraic unknowns? You don't necessarily have to be able to do all of these things right this instant, but you should at the very least recognize most of them and be able to do them with just a very short review:

- What are the two values of α that solve:

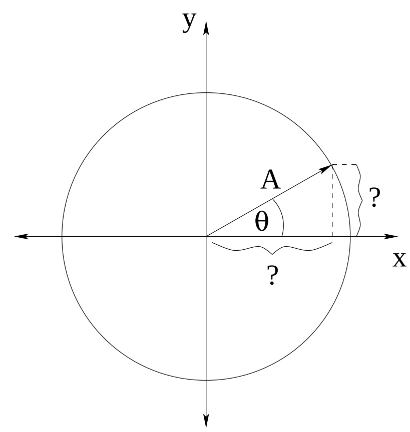
$$\alpha^2 + \frac{R}{L}\alpha + \frac{1}{LC} = 0?$$

- What is:

$$Q(r) = \frac{\rho_0}{R} 4\pi \int_0^r r'^3 dr'?$$

- What is:

$$\frac{d\cos(\omega t + \delta)}{dt}?$$



- What are the x and y components of a vector of length A that makes an angle of θ with the positive x axis (proceeding, as usual, counterclockwise for positive θ)?

- What is the sum of the two vectors $\vec{\boldsymbol{A}} = A_x\hat{\boldsymbol{x}} + A_y\hat{\boldsymbol{y}}$ and $\vec{\boldsymbol{B}} = B_x\hat{\boldsymbol{x}} + B_y\hat{\boldsymbol{y}}$?

- What is the inner/dot product of the two vectors $\vec{\boldsymbol{A}} = A_x\hat{\boldsymbol{x}} + A_y\hat{\boldsymbol{y}}$ and $\vec{\boldsymbol{B}} = B_x\hat{\boldsymbol{x}} + B_y\hat{\boldsymbol{y}}$?

- What is the cross product of the two vectors $\vec{\boldsymbol{r}} = r_x\hat{\boldsymbol{x}}$ and $\vec{\boldsymbol{F}} = F_y\hat{\boldsymbol{y}}$ (magnitude and direction)?

If *all* of these items are unfamiliar – you don't remember the quadratic formula (needed to solve the first one), can't integrate $x^n dx$ (needed to solve the second one), don't recall how to differentiate a sine or cosine function, don't recall your basic trigonometry so that you can't find the components of a vector from its length and angle or vice versa, and don't recall what the dot or cross product of two vectors are, then you are going to have to *add* to the burden of learning physics per se the burden of learning, or re-learning, all of the basic mathematics that would have permitted you to answer these questions easily.

Here are the answers, see if this jogs your memory:

- Here are the two roots, found with the quadratic formula:

$$\alpha_\pm = \frac{-\frac{R}{L} \pm \sqrt{\left(\frac{R}{L}\right)^2 - \frac{4}{LC}}}{2} = -\frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$$

- $$Q(r) = \frac{\rho_0}{R} 4\pi \int_0^r r'^3 dr' = \frac{\rho_0}{R} 4\pi \left. \frac{r'^4}{4} \right|_0^r = \frac{\rho_0 \pi r^4}{R}$$

- $$\frac{d\cos(\omega t + \delta)}{dt} = -\omega \sin(\omega t + \delta)$$

- $$A_x = A\cos(\theta) \qquad\qquad A_y = A\sin(\theta)$$

- $$\vec{\boldsymbol{A}} + \vec{\boldsymbol{B}} = (A_x + B_x)\hat{\boldsymbol{x}} + (A_y + B_y)\hat{\boldsymbol{y}}$$

- $$\vec{\boldsymbol{A}} \cdot \vec{\boldsymbol{B}} = A_x B_x + A_y B_y$$

- $$\vec{\boldsymbol{r}} \times \vec{\boldsymbol{F}} = r_x \hat{\boldsymbol{x}} \times F_y \hat{\boldsymbol{y}} = r_x F_y (\hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}}) = r_x F_y \hat{\boldsymbol{z}}$$

My strong *advice* to you, if you are now feeling the cold icy grip of panic because in fact you are signed up for physics using this book and you couldn't answer *any* of these questions and don't even *recognize* the answers once you see them, is to seek out the course instructor and review your math skills with him or her to see if, in fact, it is advisable for you to take physics at this time or rather should wait and strengthen your math skills first. You can, and will, learn a lot of math while taking physics and that is actually part of the point of taking it! If you are *too* weak going into it, though, it will cost you some misery and hard work and some of the grade you might have gotten with better preparation ahead of time.

So, what if you *could* do at least *some* of these short problems and can remember once learning/knowing the tools, like the Quadratic Formula, that you were *supposed* to use to solve them? Suppose you are pretty sure that – given a chance and resource to help you out – you can do some review and they'll all be fresh once again in time to keep up with the physics and still do well in the course? What if you have no *choice* but to take physics now, and are just going to have to do your best and relearn the math as required along the way? What if you did in fact understand math pretty well once upon a time and are sure it won't be *much* of an obstacle, but you really would like a review, a summary, a listing of the things you need to know someplace handy so you can instantly look them up as you struggle with the problems that uses the math it contains? What if you are (or were) *really good* at math, but want to be able to look at derivations or reread explanations to bring stuff you learned right back to your fingertips once again?

Hmmm, that set of questions spans the set of student math abilities from the near-tyro to the near-expert. In my experience, *everybody* but the most mathematically gifted students can probably benefit from having a math review handy while taking this course. For all of you, then, I provide the following free book online:

Mathematics for Introductory Physics

It is located here:

http://www.phy.duke.edu/~rgb/Class/math_for_intro_physics.php

It is a work in progress, and is quite possibly still somewhat incomplete, but it should help you with a *lot* of what you are missing or need to review, and if you let *me* know what you are missing that you didn't find there, I can work to add it!

I would strongly advise all students of introductory physics (any semester) to visit this site *right now* and bookmark it or download the PDF, and to visit the site from time to time to see if I've posted an update. It is on my back burner, so to speak, until I finish the actual physics texts themselves that I'm working on, but I will still add things to them as motivated by my own needs teaching courses using this series of books.

1.6: Summary

That's enough preliminary stuff. At this point, if you've read all of this "week"'s material and vowed to adopt the method of three passes in all of your homework efforts, if you've bookmarked the math help or downloaded it to your personal ebook viewer or computer, if you've realized that your brain is actually something that you can *help and enhance* in various ways as you try to learn things, then my purpose is well-served and you are as well-prepared as you can be to tackle physics.

Homework for Week 0

Problem 1.

Skim read this entire section (Week 0: How to Learn Physics), then read it like a novel, front to back. Think about the connection between engagement and learning and how important it is to try to have *fun* in a physics course. Write a short essay (say, three paragraphs) describing at least one time in the past where you were extremely engaged in a course you were taking, had lots of fun in the class, and had a really great learning experience.

Problem 2.

Skim-read the entire content of ***Mathematics for Introductory Physics*** (linked above). Identify things that it covers that you don't remember or don't understand. Pick one and learn it.

Problem 3.

Apply the *Method of Three Passes* to *this* homework assignment. You can either write three short essays or revise your one essay three times, trying to improve it and enhance it each time for the first problem, and review both the original topic and any additional topics you don't remember in the math review problem. On the *last* pass, write a short (two paragraph) essay on whether or not you found multiple passes to be effective in helping you remember the content.

Note well: You may well have found the content *boring* on the third pass because it was so familiar to you, but that's not a bad thing. If you learn physics so thoroughly that its laws become *boring*, not because they confuse you and you'd rather play World of Warcraft but because you know them so well that reviewing them isn't adding anything to your understanding, well *damn* you'll do well on the exams testing the concept, won't you?

II: Elementary Mechanics

OK, so now you are ready to learn physics. Your math skills are buffed and honed, you've practiced the method of three passes, you understand that success depends on your full engagement and a certain amount of hard work. In case you missed the previous section (or are unused to actually reading a math-y textbook instead of minimally skimming it to extract just enough "stuff" to be able to do the homework) I usually review its content on the first day of class at the same time I review the syllabus and set down the class rules and grading scheme that I will use.

It's time to embark upon the actual week by week, day by day progress through the course material. For maximal ease of use for you the student and (one hopes) your instructor whether or not that instructor is me, the course is designed to cover *one chapter per week-equivalent*, whether or not the chapter is broken up into a day and a half of lecture (summer school), an hour a day (MWF), or an hour and a half a day (TTh) in a semester based scheme. To emphasize this preferred rhythm, each chapter will be referred to by the *week* it would normally be covered in my own semester-long course.

A week's work in all cases covers just about exactly one "topic" in the course. A very few are spread out over two weeks; one or two compress two related topics into one week, but in all cases the *homework* is assigned on a weekly rhythm to give you ample opportunity to use the *method of three passes* described in the first part of the book, culminating in an expected 2-3 hour *recitation* where you should go over the assigned homework *in a group* of three to six students, with a mentor handy to help you where you get stuck, with a goal of *getting all of the homework perfectly correct by the end of recitation*.

That is, at the end of a week plus its recitation, you *should* be able to do *all* of the week's homework, *perfectly*, and *without looking or outside help*. You will usually *need* all three passes, the last one working in a group, *plus* the mentored recitation to achieve this degree of competence! But without it, surely the entire process is a waste of time. Just *finishing* the homework is not enough, the whole point of the homework is to help you learn the material and it is the latter that is the real goal of the activity not the mere completion of a task.

However, *if* you do this – attempt to really master the material – you are almost certain to do well on a quiz that terminates the recitation period, and you will be very likely to *retain* the material and not have to "cram" it in again for the hour exams and/or final exam later in the course. Once you achieve *understanding* and reinforce it with a fair bit of repetition and practice, most students will naturally transform this experience into remarkably deep and permanent learning.

Note well that each week is organized for *maximal ease of learning* with the week/chapter review *first*. Try to *always look at this review before lecture* even if you skip reading the chapter itself until later, when you start your homework. Skimming the whole week/chapter guided by this summary before lecture is, of course, better still. It is a "first pass" that can often make lecture much easier to follow and help free you from the tyranny of note-taking as you only need to note *differences* in the presentation from this text and perhaps the answers to *questions* that helped you understand something during the discussion. Then read or skim it again right before each homework pass.

Week 1: Newton's Laws

1.7: Summary

- Physics is a *language* – in particular the language of a certain kind of *worldview*. For philosophically inclined students who wish to read more deeply on this, I include links to terms that provide background for this point of view.
  - Wikipedia: http://www.wikipedia.org/wiki/Worldview
  - Wikipedia: http://www.wikipedia.org/wiki/Semantics
  - Wikipedia: http://www.wikipedia.org/wiki/Ontology

  Mathematics is the natural language and logical language of physics, not for any particularly deep reason but because it *works*. The components of the semantic language of physics are thus generally expressed as mathematical or logical *coordinates*, and the semantic expressions themselves are generally mathematical/algebraic *laws*.

- **Coordinates** are the *fundamental adjectival modifiers* corresponding to the differentiating properties of "things" (nouns) in the real Universe, where the term fundamental can also be thought of as meaning *irreducible* – adjectival properties that cannot be readily be expressed in terms of or derived from other adjectival properties of a given object/thing. See:
  - Wikipedia: http://www.wikipedia.org/wiki/Coordinate System

- **Units.** Physical coordinates are basically mathematical numbers with units (or can be so considered even when they are discrete non-ordinal sets). In this class we will consistently and universally use *Standard International* (SI) units unless otherwise noted. Initially, the irreducible units we will need are:

  a) *meters* – the SI units of *length*

  b) *seconds* – the SI units of *time*

  c) *kilograms* – the SI units of *mass*

  All other units for at least a short while will be expressed in terms of these three, for example units of *velocity* will be meters per second, units of *force* will be kilogram-meters per second squared. We will often give names to some of these combinations, such as the SI units of force:

$$1 \text{ Newton} = \frac{\text{kg-m}}{\text{sec}^2}$$

Later you will learn of other irreducible coordinates that describe elementary particles or extended macroscopic objects in physics such as electrical charge, as well as additional derivative quantities such as energy, momentum, angular momentum, electrical current, and more.

- **Laws of Nature** are essentially *mathematical postulates* that permit us to understand natural phenomena and *make predictions* concerning the time evolution or static relations between the coordinates associated with objects in nature that are consistent mathematical theorems of the postulates. These predictions can then be compared to *experimental observation* and, if they are consistent (uniformly successful) we *increase our degree of belief in them*. If they are inconsistent with observations, we *decrease* our degree of belief in them, and seek alternative or modified postulates that work better[28].

  The entire body of human scientific knowledge is the more or less successful outcome of this process. This body is not fixed – indeed it is *constantly changing* because it is an *adaptive* process, one that self-corrects whenever observation and prediction fail to correspond or when new evidence spurs insight (sometimes revolutionary insight!)

  Newton's Laws built on top of the analytic geometry of Omar Khayyam[29] and Rene Descartes[30] (as the basis for at least the abstract spatial coordinates of things) are the *dynamical principle* that proved *successful* at predicting the outcome of many, many everyday experiences and experiments as well as cosmological observations in the late 1600's and early 1700's all the way up to the mid-19th century[31]. When combined with associated empirical *force laws* they form the basis of the physics you will learn in this course.

- **Newton's Laws**:

  a) **Law of Inertia:** Objects at rest or in uniform motion (at a constant velocity) in an *inertial reference frame* remain so unless acted upon by an unbalanced (net) force.

[28] Students of philosophy or science who *really* want to read something cool and learn about the fundamental basis of our knowledge of reality are encouraged to read e.g. Richard Cox's *The Algebra of Probable Reason* or E. T. Jaynes' book *Probability Theory: The Logic of Science*. These two related works *quantify* how science is not (as some might think) absolute truth or certain knowledge, but rather *the best set of things to believe* based on our overall experience of the world, that is to say, "the evidence".

[29] Wikipedia: http://www.wikipedia.org/wiki/Omar_Khayyam. It has been argued – with some reason – that Omar Khayyam laid the foundations for Rene Descartes eventual invention of analytic geometry. Without rendering a personal judgement on the issue, I think it is pretty reasonable to honor him here as he *was* almost certainly the first person to mix methods of algebra and geometry and use them together in the solution of a wide range of problems.

[30] Wikipedia: http://www.wikipedia.org/wiki/Rene_Descartes.

[31] Although they failed in the late 19th and early 20th centuries, to be superceded by relativistic quantum mechanics. Basically, everything we learn in this course is *wrong*, but it nevertheless *works damn well* to describe the world of macroscopic, slowly moving objects of our everyday experience.

We can write this algebraically[32] as

$$\sum_i \vec{\boldsymbol{F}}_i = 0 = m\vec{\boldsymbol{a}} = m\frac{d\vec{\boldsymbol{v}}}{dt} \quad \Rightarrow \quad \vec{\boldsymbol{v}} = \text{constant vector} \tag{1.1}$$

b) **Law of Dynamics:** The net force applied to an object is directly proportional to its acceleration in an *inertial reference frame*. The constant of proportionality is called the **mass** of the object. We write this algebraically as:

$$\vec{\boldsymbol{F}} = \sum_i \vec{\boldsymbol{F}}_i = m\vec{\boldsymbol{a}} = \frac{d(m\vec{\boldsymbol{v}})}{dt} = \frac{d\vec{\boldsymbol{p}}}{dt} \tag{1.2}$$

where we introduce the *momentum* of a particle, $\vec{\boldsymbol{p}} = m\vec{\boldsymbol{v}}$, in the last way of writing it.

c) **Law of Reaction:** If object B exerts a force $\vec{\boldsymbol{F}}_{AB}$ on object A *along a line connecting the two objects*, then object A exerts an equal and opposite reaction force of $\vec{\boldsymbol{F}}_{BA} = -\vec{\boldsymbol{F}}_{AB}$ on object B. We write this algebraically as:

$$\vec{\boldsymbol{F}}_{ij} = -\vec{\boldsymbol{F}}_{ji} \tag{1.3}$$

$$\text{(or)} \quad \sum_{i,j} \vec{\boldsymbol{F}}_{ij} = 0 \tag{1.4}$$

(the latter form means that the sum of all *internal* forces between particles in any closed system of particles cancel).

- **An inertial reference frame** is a ***spatial coordinate system plus an independent time coordinate*** that is either at rest or moving at a constant speed, a *non-accelerating* set of coordinates that can be used to describe the locations of real objects as a function of time. However, this definition is inadequate, because acceleration itself is defined only *relative to a frame*. This leaves us with a problem: non-accelerating relative to *what frame*? We have to identify at least one inertial reference frame before we can talk about other frames that do not accelerate relative to *it*.

  The way we can identify *any* inertial reference frame is as follows. As you will see, ***all actual physical forces in the laws of nature*** are ***interaction*** laws. For any observed force pushing an object, there exists *another object* somewhere that is doing the pushing, a "Newton's Third Law partner". No force exists in isolation, and no object can exert a force on itself. A signature of a *non*-inertial reference frame is that within it, there exists an observed "pseudoforce" that arises in Newton's Second Law due to the acceleration of the frame. This pseudoforce ***has no Newton's Third Law partner!*** A *consistent* definition of an inertial reference frame is therefore:

  > **Inertial Reference Frame**: ***Any frame where all observed forces that occur in all statements of Newton's Second Law for all particles are pairwise interactions between particles.*** In other words, there are no forces that act on any particle in complete isolation from and independent of the other particles in the system

[32]For students who are still feeling very shaky about their algebra and notation, let me remind you that $\sum_i \vec{\boldsymbol{F}}_i$ stands for "The sum over i of all force $\vec{\boldsymbol{F}}_i$", or $\vec{\boldsymbol{F}}_1 + \vec{\boldsymbol{F}}_2 + \vec{\boldsymbol{F}}_3 + ...$ We will often use $\sum$ as shorthand for summing over a list of similar objects or components or parts of a whole.

.

In physics one has considerable leeway when it comes to choosing the (inertial) coordinate frame to be used to solve a problem – some lead to much simpler solutions than others!

- **Forces of Nature** (weakest to strongest):

  a) Gravity
  b) Weak Nuclear
  c) Electromagnetic
  d) Strong Nuclear

  It is possible that there are more forces of nature waiting to be discovered. Because physics is not a *dogma*, this presents no real problem. If they are, we'll simply give the discoverer a Nobel Prize and add their name to the "pantheon" of great physicists, add the force to the list above, and move on. Science, as noted, is self-correcting.

- **Force is a vector**. For each force rule below we therefore need *both* a formula or rule for the magnitude of the force (which we may have to compute in the solution to a problem – in the case of forces of constraint such as the normal force (see below) we will *usually* have to do so) and a way of determining or specifying the *direction* of the force. Often this direction will be obvious and in corresponence with experience and mere common sense – strings pull, solid surfaces push, gravity points down and not up. Other times it will be more complicated or geometric and (along with the magnitude) may vary with position and time.

- **Force Rules** The following set of force rules will be used both in this chapter and throughout this course. All of these rules can be derived or understood (with some effort) from the forces of nature, that is to say from "elementary" natural laws.

  a) **Gravity** (near the surface of the earth):

  $$F_g = mg \tag{1.5}$$

  The direction of this force is *down*, so one could write this in vector form as $\vec{\boldsymbol{F}}_g = -mg\hat{\boldsymbol{y}}$ in a coordinate system such that up is the $+y$ direction. This rule follows from Newton's Law of Gravitation, the elementary law of nature in the list above, evaluated "near" the surface of the earth where it is approximately constant.

  b) **The Spring** (Hooke's Law) in one dimension:

  $$F_x = -k\Delta x \tag{1.6}$$

  This force is directed back to the equilibrium point of unstretched spring, in the *opposite direction* to Δx, the displacement of the mass on the spring from equilibrium. This rule arises from the primarily electrostatic forces holding the atoms or molecules of the spring material together, which tend to linearly oppose *small* forces that pull them apart or push them together (for reasons we will understand in some detail later).

c) The **Normal Force**:

$$F_\perp = N \tag{1.7}$$

This points perpendicular and away from solid surface, magnitude sufficient to oppose the force of contact *whatever it might be!* This is an example of a *force of constraint* – a force whose magnitude is determined by the *constraint* that one solid object cannot generally interpenetrate another solid object, so that the solid surfaces exert whatever force is needed to prevent it (up to the point where the "solid" property itself fails). The physical basis is once again the electrostatic molecular forces holding the solid object together, and *microscopically* the surface deforms, however slightly, more or less like a spring.

d) **Tension** in an Acme (massless, unstretchable, unbreakable) string:

$$F_s = T \tag{1.8}$$

This force simply transmits an *attractive* force between two objects on opposite ends of the string, in the directions of the taut string at the points of contact. It is another constraint force with no fixed value. Physically, the string is like a spring once again – it microscopically is made of bound atoms or molecules that pull ever so slightly apart when the string is stretched until the restoring force balances the applied force.

e) **Static Friction**

$$f_s \leq \mu_s N \tag{1.9}$$

(directed opposite towards net force parallel to surface to contact). This is another force of constraint, as large as it needs to be to keep the object in question travelling at the same speed as the surface it is in contact with, up to the *maximum* value static friction can exert before the object starts to slide. This force arises from mechanical interlocking at the microscopic level plus the electrostatic molecular forces that hold the surfaces themselves together.

f) **Kinetic Friction**

$$f_k = \mu_k N \tag{1.10}$$

(opposite to direction of relative sliding motion of surfaces and parallel to surface of contact). This force *does* have a fixed value when the right conditions (sliding) hold. This force arises from the forming and breaking of microscopic adhesive bonds between atoms on the surfaces plus some mechanical linkage between the small irregularities on the surfaces.

g) **Fluid Forces, Pressure**: A fluid in contact with a solid surface (or anything else) in general exerts a force on that surface that is related to the *pressure* of the fluid:

$$F_P = PA \tag{1.11}$$

which you should read as "the force exerted by the fluid on the surface is the pressure in the fluid times the area of the surface". If the pressure varies or the surface is curved one may have to use calculus to add up a total force. In general the direction of the force exerted is *perpendicular* to the surface. An object at rest in a fluid often has balanced forces due to pressure. The force arises from the molecules in the fluid literally bouncing off of the surface of the object, transferring momentum (and exerting an average force) as they do so. We will study this in some detail and will even derive a kinetic model for a gas that is in good agreement with real gases.

h) **Drag Forces**:

$$F_d = -bv^n \tag{1.12}$$

(directed opposite to relative velocity of motion through fluid, n usually between 1 (low velocity) and 2 (high velocity). This force also has a determined value, although one that depends in detail on the motion of the object. It arises first because the surface of an object moving through a fluid is literally bouncing fluid particles off in the leading direction while moving away from particles in the trailing direction so that there is a differential pressure on the two surfaces, second from "friction" between the fluid particles and the surface.

The first week summary would not be complete without some sort of reference to *methodologies of problem solving* using Newton's Laws and the force laws or rules above. The following rubric should be of great use to you as you go about solving *any* of the problems you will encounter in this course, although we will modify it slightly when we treat e.g. energy instead of force, torque instead of force, and so on.

Dynamical Recipe for Newton's Second Law

a) *Draw* a good picture of what is going on. In general you should probably do this even if one has been provided for you – visualization is key to success in physics.

b) On your drawing (or on a second one) decorate the objects with all of the *forces* that act on them, creating a *free body diagram* for the forces on each object. It is sometimes useful to draw pictures of each object in isolation with ***just*** the forces acting on that one object connected to it, although for simple problems this is not always necessary. Either way your diagram should be clearly drawn and labelled.

c) Choose a suitable coordinate system for the problem. This coordinate system need not be cartesian (x, y, z). We sometimes need separate coordinates for each mass (with a relation between them) or will even find it useful to "wrap around a corner" (following a string around a pulley, for example) in some problems

d) *Decompose* the forces acting on each object into their components in the (orthogonal) coordinate frame(s) you chose, using trigonometry and geometry.

e) *Write* Newton's Second Law for each object (summing the forces and setting the result to $m_i\vec{a}_i$ for each – ith – object for each dimension) and algebraically rearrange it into (vector) differential equations of motion (practically speaking, this means solving for or isolating the *acceleration* $\vec{a}_i = \frac{d^2\vec{x}_i}{dt^2}$ of the particles in the equations of motion).

f) *Solve* the independent 1 dimensional systems for each of the independent orthogonal coordinates chosen, plus any coordinate system constraints or relations. In many problems the *constraints* will eliminate one or more degrees of freedom from consideration, especially if you have chosen your cooordinates wisely (for example, ensuring that one coordinate points in the direction of a known component of the acceleration such as 0 or $\Omega^2 r$).

   Note that in most nontrivial cases, these solutions will have to be *simultaneous* solutions, obtained by e.g. algebraic substitution or elimination.

g) *Reconstruct* the multidimensional trajectory by adding the vector components thus obtained back up (for a common independent variable, time). In some cases you may skip straight ahead to other known kinematic relations useful in solving the problem.

h) *Answer* algebraically any questions requested concerning the resultant trajectory, using kinematic relations as needed.

Some parts of this rubric will require *experience and common sense* to implement correctly for any given kind of problem. That is why homework is so critically important! We want to make solving the problems (and the conceptual understanding of the underlying physics) *easy*, and they will only get to be easy with practice followed by a certain amount of meditation and reflection, practice followed by reflection, iterate until the light dawns...

1.1: Introduction: A Bit of History and Philosophy

In order to learn physics, we have to start by learning just what physics itself *is*. It's easy to confuse physics as a subject to be learned and physics as a synonym for "fundamental reality". To avoid this confusion, let's think about just what physics is (and isn't!), where physics came from, and why it is important and useful for us to learn it.

Physics is the core component of the ***scientific worldview***[33] , which is the *defensibly best epistemology/ontology*[34] ,[35] invented thus far by the human species. It is important to differentiate a couple of things here. As best as we can tell using our senses, there is an objectively real external Universe that "we" (you if you are reading this, I if I'm still alive, as I am while I write this) live "in" – or better, as a part of. To avoid all confusion, I'm going to define that Universe as "the set of everything that has the objective property of actual existence" – anywhere, any time, in any dimensionality. If it is real, a part of the fundamental ontology, it is a part of the Universe, real independent of whether there is anything that can be considered *knowledge* of, or better, *beliefs* about, that Universe.

This Universe is in some fundamental sense distinct from the *epistemological knowledge* of it. This is obvious in many ways. For one, our knowledge/beliefs about it can be *incorrect*, but the Universe itself is beyond "correct" or "incorrect" – it just "is". In some deep sense, our correct beliefs are those that are (as it turns out) in actual agreement with this external reality, an agreement or correspondence we are never able to *prove* with *absolute certainty* but that is nevertheless sufficiently compelling to be considered to be "proven true" to some working approximation of the meaning of the word truth, one perhaps best expressed as being ***very likely*** to be true by virtue of being in the overall best agreement with observations of reality (of all proposed models of reality) while also remaining consistent with the entire network of best, evidence supported beliefs obtained ***so far***.

Physics, then, is – like all knowledge – nothing more than a working set of strongly held *beliefs* about the fundamental nature of being of the Universe itself. This is *why* it is literally the most fundamental level of the *scientific worldview*, the scientific epistemology as a work in progress: the set of mutually consistent, empirically supported beliefs that have the greatest predictive and explanatory power of any set of reasonably mutually consistent proposed beliefs and that are therefore accepted in consensus as most likely to be true – so far[36].

It's important to remember this as you start to learn physics at the introductory level, because pretty much everything I'm going to try to teach you about physics in the standard two semester college course of physics (with calculus) is – technically, in very well known ways – *incorrect*. Specifically, it is *not* the best, most accurate and consistent set of beliefs concern-

---

[33]Wikipedia: http://www.wikipedia.org/wiki/Worldview. In a nutshell, a worldview is the entirety of your belief system and linguistic symbolic semantic system relating to "how the world you appear to inhabit works".

[34]Wikipedia: http://www.wikipedia.org/wiki/Epistemology. In a nutshell, epistemology is "how we know things".

[35]Wikipedia: http://www.wikipedia.org/wiki/Ontology. In the *same* nutshell (crowded in here), ontology is a set of concepts in a domain as well as the relationships between them. In this context the physical ontology is "the nature of being of the Universe itself". A worldview is an ontology although the two terms aren't *quite* synonymous.

[36]If, by the end of studying this text or taking the associated course, you learn and conceptually understand *just this one thing* I will consider the course a success in your personal case. Once you grasp this, supernaturalism becomes literally self-contradictory and you have a *chance* to understand the world around as best we all, working together, can consistently explain it.

ing the Universe we've obtained so far, and it will *not* have the widest possible predictive and explanatory power. In some – indeed nearly all – of the cases we will apply physics to, the errors in what we will learn will be indetectably small and irrelevant, much smaller than a host of other factors that affect the "resolution" and accuracy of our observations of nature. In a few, very famous cases, they will not.

Specifically, we will learn classical, non-relativistic physics – almost entirely the physics of the eighteenth and early nineteenth century. By the late nineteenth and early twentieth century, classical physics was well on the way to being superceded by relativistic quantum physics – physics that smoothly transforms *in*to classical non-relativistic physics in the appropriate scaling regime, but that works much better when studying the very small or the very fast. These are the regimes and theories that you will eventually study if you move on to modern physics, but even there, even now, our understanding of e.g. relativistic quantum field theory is solidly grounded in the *concepts first learned* in classical mechanics

Physics (like all of science) is also *incomplete*. You may get the feeling in this class that what we are learning is sufficient to explain *everything*, or would be if we had more/better knowledge of the state of everything and included relativity theory and quantum mechanics. This is not true. We run into multiple barriers – some logical/mathematical (like Godël's incompleteness theorems), some due to fundamental limits on what we can observe of nature at both the largest and smallest length and time scales, some due to economics, some due to mere distance and obstruction, some (to be frank) due to the limitations of our intelligence and imagination.

After all, we *know* that we will quite literally *never* be able to observe any significant fraction of the Universe whose existence we can legitimately infer from observations we *can* make, let alone build a successful *detailed* model of in good agreement with observation. We can only "see" a kind of time-lagged "slice" through the space-time of the Universe leading back to the Big Bang, and can fairly confidently state that the actual Universe is at least *hundreds of times larger* than the portion physics *itself* limits us to seeing in the fourteen-billion-odd years since the Big Bang. The best we can hope to do is build a *working model* that explains everything we *can* see in general terms, predict the results of experiments that *are* accessible to us, and acknowledge that 99.9999...% (really, much ***much*** more) of the *details* of the state of the Universe are literally impossible for us to know.

Let's then leave off trying to accomplish "perfect knowledge" and concentrate on the practical. Consider for the moment some of the *advantages* of learning physics as human beings living in a complex ecology. After all, that's what we are. Does understanding physics give us a survival/reproduction edge in the ongoing dance of evolution? Let's think about how living organisms interact with their environment.

The most elementary objects in the tree of life that *I* know of[37] – prions – are misfolded proteins only slightly removed from pure chemistry. Viruses also simply hijack (at a higher DNA/RNA level) protein synthesis processes in cells. The most elementary single-celled "living" organisms have nothing beyond the most elementary chemical "memory'.

In one entire phylogenetic tree, plants as a rule react to immediate stimuli in very limited, essentially "automatic" ways. They have no detectable "memory" of complex phenomena or

[37] Which isn't saying much as I'm not a biologist, be warned...

the attendant ability to think and do not exhibit "behavior" beyond fairly simple, usually slow, tropisms[38] .

In the other tree, simple animals exhibit first "real time behavior" and then (as they become multicellular organisms) increasingly behave as if they have a *memory* of their personal (and species!) history that they can use to guide their behavior to optimize their biological imperatives of food, safety, and reproduction.

At some point animals start to exhibit "intelligence" – the ability to transform memory into *working models* for their environment that can *predict its evolution in time!* Again, they (can) use this model to more successfully accomplish their evolutionary goals of survival and reproduction.

Humans are unique in our ability to transform memory and experience into *symbols* and create *symbolic models* of the world that for better or worse help them to survive/reproduce more successfully, at least as far as our observations confined to planet Earth are concerned. The earliest models recognizably "human" cultures developed were extremely crude and anthropomorphic. Confronted with a complex world and little understanding or control of it, humans invented supernatural agencies galore to explain everyday occurrences[39].

Lacking the cognitive/epistemological tools to do much better, they codified *social* rules that hid most of the detail but worked. Times to plant and harvest crops, rules for social order, explanations for disease, death, lightning, drought, and more all incorporated into a mix of custom, law, and divine/religious rules. The many gods filled the many gaps in our knowledge and understanding.

Roughly 2500 years ago, several cultures started to do better. The ancient Greeks, in particular, invented "secular" philosophy (separated from their religion), mathematics, logic, codified ethics, and – the first theories of *physics*. They were quite successful and incorporated empiricism – they learned that the Earth was round and that it orbited the Sun and not the other way around, for example – but they lacked certain tools and failed to pass a critical methological barrier!

These advances were not limited to just the Greeks. In various historical epochs, pieces of science and mathematics were developed in India, in China, in greater Arabia, in the Americas. But in all cases there were *missing* pieces and the crucial ideas that led to the modern era were not discovered until roughly 2000 more years after the invention of non-religious philosophy.

By the Middle Ages, prosperity, education, and other factors conspired to give birth to the idea that the best way to determine how the world works is to *build symbolic models* and *test the models empirically against the observable world!* (Names: Roger Bacon, Bernardino Telesio, Giordano Bruno, Francis Bacon...)

Some of the key advances (many of which we will study in this textbook):

- The Copernican heliocentric theory
- Galileo's observational support of this model.

[38]Wikipedia: http://www.wikipedia.org/wiki/Plant_Memory. As will often be the case in my sweeping statements in the course, this is true only to the extent that one quibbles about the meaning of "complex". But I already warned you that I'm presenting an oversimplified view of the world.

[39]Some of this continues to this very day.

- Tycho Brahe's observational data of planetary orbits.
- Johannes Kepler's analysis of Brahe's data.
- Omar Khayyam, poet-mathematician, was the inventor of the first versions and proofs in algebraic geometry.
- René Descartes, philosopher-mathematician, invented **analytic geometry** (almost certainly heavily based on Khayyam's work).

All of these set the empirical stage for Isaac Newton to co-invent the *physics* we will learn and the *calculus* it is based on. Descartes analytical geometry was perhaps the most critical component of Newton's invention of calculus and we will therefore start our own learning process by exploring the ideas of **spatial coordinate systems** (especially Cartesian coordinates – standard x, y, z coordinate *frames*, but also including polar coordinates and more later on) used to describe the *motion of something!*

Because many of these advances heretically challenged established religious pronouncements, much of this early work was banned. Galileo was famously prosecuted for heresy and spent the last years of his life under house arrest, unable to further contribute to the physics he came close to inventing. Less famously ***Giordano Bruno***, a priest who dared to openly hypothesize that the stars in the sky were actual suns like our own, orbited by planets like our own, that might well support life – like our own – was ***burned at the stake*** for his temerity. This violent reaction to new ideas that contradicted the biblical worldview sent a strong message to those that dared dream – and model – alternatives.

The Newtonian model that ultimately prevailed was by no means obvious, and was iconoclastic in many ways. One reason students often have a hard time learning Newtonian physics is that they have to first *unlearn* their already prevailing "natural" worldview of physics, which dates all the way back to Aristotle. In a nutshell and in very general terms (skipping all the "nature is a source or cause of being moved and of being at rest" as primary attributes, see Aristotle's *Physica*, book II) Aristotle considered motion or the lack thereof of an object to be innate properties of materials, according to their proportion of the big four basic elements: Earth, Air, Fire and Water. He extended the idea of the moving and the immovable to cosmology and to his Metaphysics as well.

In this primitive view of things, the observation that (most) physical objects (being "Earth") set in motion slow down is translated into the notion that their natural state is to be at rest, and that one has to add *something* from one of the other essences to produce a state of uniform motion. This was not an unreasonable hypothesis; a great deal of a person's everyday experience (then and now) is consistent with this. When we pick something up and throw it, it moves for a time and bounces, rolls, slides to a halt. We need to press down on the accelerator of a car to keep it going, adding something from the "fire" burning in the motor to the "earth" of the body of the car. Even our selves seem to run on "something" that goes away when we die.

Unfortunately, it is completely and totally wrong. Indeed, it is almost precisely Newton's first law stated *backwards*. It is very likely that the *reason* Newton wrote down his first law (which is otherwise a trivial consequence of his second law) was to directly confront the error of Aristotle, to force the philosophers of his day to confront the fact that his (Newton's) theory of physics was *irreconcilable* with that of Aristotle, and that (since his actually *worked* to make precise predictions of nearly any kind of classical motion that were in good agreement with observation and experiments designed to test it) Aristotle's physics was almost certainly *wrong*. Or at any rate, wronger than Newton's.

Newton's discoveries were a core component of the Enlightment, a period of a few hundred years in which Europe went from a state of almost slavish, church-endorsed belief in the infallibility and correctness of the Aristotelian worldview to a state where humans, for the first time in history, let nature *speak for itself* by using a *consistent framework* to listen to what nature had to say[40]. Aristotle lost, but his *ideas* are slow to die *because* they closely parallel everyday experience. The *first* chore for any serious student of physics is thus to unlearn this Aristotelian view of things[41].

This is not, unfortunately, an abstract problem. It is very concrete and very current. Because I have an online physics textbook, and because physics is in some very fundamental sense the "magic" that makes the world work, I not infrequently am contacted by individuals who do *not* understand the material covered in this textbook, who do *not* want to do the very hard work required to master it, but who still want to be "magicians". So they invent their *own* version of the magic, usually altering the *mathematically precise* meanings of things like "force", "work", "energy" to be something else altogether that *they* think that they understand but that, when examined closely, usually are dimensionally or conceptually inconsistent and mean nothing at all.

Usually their "new" physics is in fact a reversion to the physics of Aristotle. They recreate the magic of earth and air, fire and water, a magic where things slow down unless fire (energy) is added to sustain their motion or where it can be extracted from invisible an inexhaustible resources, a world where the *mathematical* relations between work and energy and force and acceleration do not hold. A world, in short, that violates a huge, vast, truly stupdendous body of accumulated experimental evidence including the very evidence that you yourselves will very likely observe in person in the physics labs associated with this course. A world in which things like perpetual motion machines are possible, where free lunches abound, and where humble dilettantes can be crowned "the next Einstein" without having a solid understanding of algebra, geometry, advanced calculus, or the physics that are *just the ante* for playing the modern game of physics.

This world does not exist. Seriously, it is a fantasy, and a *very dangerous one*, one that threatens modern civilization itself. One of the *most important* reasons you are taking this course, *whatever* your long term dreams and aspirations are professionally, is to come to fully and deeply understand this. You will come to understand the magic of science, at the same time you learn to reject the notion that science *is* magic or vice versa.

---

[40] Students who like to read historical fiction will doubtless enjoy Neal Stephenson's *Baroque Cycle*, a set of novels – filled with sex and violence and humor, a good read – that spans the Enlightenment and in which Newton, Liebnitz, Hooke and other luminaries to whom we owe our modern conceptualization of physics all play active parts.

[41] This is not the last chore, by the way. Physicists have long since turned time into a coordinate just like space so that how long things take depends on one's point of view, eliminated the assumption that one can measure any set of measureable quantities to arbitrary precision in an arbitrary order, replaced the determinism of mathematically precise trajectories with a funny kind of stochastic quasi-determinism, made (some) forces into just an aspect of geometry, and demonstrated a degree of mathematical structure (still incomplete, we're working on it) beyond the wildest dreams of Aristotle or his mathematical-mystic buddies, the Pythagoreans.

There is nothing wrong with this. I personally find it very comforting that the individuals that take care of my body (physicians) and who design things like jet airplanes and automobiles (engineers) share a common and consistent *Newtonian*[42] view of just how things work, and would find it very disturbing if any of them believed in magic, in gods, in fairies, in earth, air, fire and water as constituent elements, in "crystal energies", in the power of a drawn pentagram or ritually chanted words in any context whatsoever. These all represent a sort of willful *wishful thinking* on the part of the believer, a desire for things to *not* follow the rigorous mathematical rules that they appear to follow as they evolve in time, for there to be a "man behind the curtain" making things work out as they appear to do. Or sometimes an entire pantheon.

Let me be therefore be precise. In the physics we will study week by week below, the natural state of "things" (e.g. objects made of matter) will be to move uniformly. We will learn *non*-Aristotelian physics, *Newtonian* physics. It is only when things are acted on from outside by *unbalanced forces* that the motion becomes non-uniform; they will speed up or slow down. By the time we are done, you will understand how this can still lead to the damping of motion observed in everyday life, why things *do* generally slow down. In the meantime, be aware of the problem and resist applying the Aristotelian view to real physics problems, and consider, *based on the evidence and your experiences taking this course* rejecting "magic" as an acceptable component of your personal worldview unless and until it too has some sort of objective empirical support. Which would, of course, just make it part of physics!

1.2: Dynamics

Physics is the study of *dynamics*. Dynamics is the description of the actual forces of nature that, we believe, underlie the causal structure of the Universe and are responsible for its *evolution in time*. We are about to embark upon the intensive study of a simple description of nature that introduces the concept of a *force*, due to Isaac Newton. A force is considered to be the *causal agent* that produces the *effect of acceleration* in any massive object, altering its dynamic state of motion.

Newton was not the first person to attempt to describe the underlying nature of causality. Many, many others, including my favorite 'dumb philosopher', Aristotle, had attempted this. The major difference between Newton's attempt and previous ones is that Newton did not frame his as a philosophical postulate per se. Instead he formulated it as a *mathematical theory* and proposed a set of *laws* that (he hoped) precisely described the regularities of motion in nature.

In physics a law is the equivalent of a *postulated axiom* in mathematics. That is, a physical law is, like an axiom, an *assumption* about how nature operates that *not* formally provable by any means, including experience, within the theory. A physical law is thus not considered "correct" – rather we ascribe to it a "degree of belief" based on how well and consistently it describes nature in experiments designed to verify *and* falsify its correspondence.

It is important to do both. Again, interested students are are encouraged to look up Karl

[42]Newtonian or better, that is. Of course actual modern physics is non-Newtonian quantum mechanics, but this is just as non-magical and concrete and it reduces to Newtonian physics in the macroscopic world of our quotidian experience.

Popper's "Falsifiability"[43] and the older Postivism[44] . A hypothesis must successfully withstand the test of repeated, reproducible experiments that *both* seek to disprove it *and* to verify that it has predictive value in order to survive and become plausible. And even then, it could be wrong!

If a set of laws survive all the experimental tests we can think up and subject it to, we consider it *likely* that it is a good approximation to the true laws of nature; if it passes many tests but then fails others (often failing consistently at some length or time scale) then we may continue to call the postulates laws (applicable within the appropriate milieu) but recognize that they are only approximately true and that they are superceded by some more fundamental laws that are closer (at least) to being the "true laws of nature".

Newton's Laws, as it happens, are in this latter category – early postulates of physics that worked remarkably well up to a point (in a certain "classical" regime) and then failed. They are "exact" (for all practical purposes) for massive, large objects moving slowly compared to the speed of light[45] for long times such as those we encounter in the everyday world of human experience (as described by SI scale units, discussed in context as we proceed). They fail badly (as a basis for prediction) for microscopic phenomena involving short distances, small times and masses, for very strong forces, and for the laboratory description of phenomena occurring at relativistic velocities. Nevertheless, even here they survive in a distorted but still recognizable form, and *some* of the constructs they introduce to help us study dynamics still survive.

Interestingly, Newton's laws lead us to second order differential equations, and even quantum mechanics appears to be based on differential equations of second order or less. Third order and higher systems of differential equations seem to have potential problems with temporal causality (where effects always follow, or are at worst simultaneous with, their causes in time); it is part of the genius of Newton's description that it precisely and sufficiently allows for a full description of causal phenomena, even where the details of that causality turn out to be incorrect.

Incidentally, one of the other interesting features of Newton's Laws is that *Newton invented calculus* to enable him to solve the problems they described. Now you know why calculus is so essential to physics: physics was the original motivation behind the invention of calculus itself. Calculus was also (more or less simultaneously) invented in the more useful and recognizable form that we still use today by other mathematical-philosophers such as Leibnitz, and further developed by many, many people such as Gauss, Poincare, Poisson, Laplace and others. In the overwhelming majority of cases, especially in the early days, solving one or more problems in the physics that was still being invented was the motivation behind the most significant

[43]Wikipedia: http://www.wikipedia.org/wiki/Falsifiability. Popper considered the ability to in principle *disprove* a hypothesis as an essential criterion for it to have objective meaning. Students might want to purchase and read Nassim Nicholas Taleb's book *The Black Swan* to learn of the dangers and seductions of worldview-building gone awry due to insufficient skepticism or a failure to allow for the disproportionate impact of the unexpected but true anyway – such as an experiment that falsifies a conclusion that was formerly accepted as being verified.

[44]Wikipedia: http://www.wikipedia.org/wiki/Positivism. This is the correct name for "verifiability", or the ability to verify a theory as the essential criterion for it to have objective meaning. The correct modern approach in physics is to do both, following the procedure laid out by Richard Cox and E. T. Jaynes wherein propositions are never proven or disproven per se, but rather are shown to be more or less "plausible". A hypothesis in this approach can have meaning as a very implausible notion quite independent of whether or not it can be verified or falsified – yet.

[45]$c = 3 \times 10^8$ meters/second

developments in calculus and differential equation theory. This trend continues today, with physics providing an underlying structure and motivation for the development of much of the most advanced mathematics.

1.3: Coordinates

1.3.1: Kinematics: 1D Position as a Function of Time

To get started, let's see how to build a kind of "map" of actual (simple!) motion. Imagine, if you will, a car moving at a constant speed of 2 meters per second in a straight line past your house. Suppose you have a time-lapse camera that takes a series of pictures, one per second as it moves along.

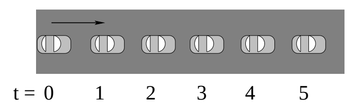


Figure 1.1: A labelled "time-lapse series" of pictures of the location of the car on the street.

We'd like to *understand* this motion, that is, to ***build a model*** of it that will eventually let us *predict* what a similar car will do if it drives down the street in the same manner. First we visualize the pictures as in figure 1.1:

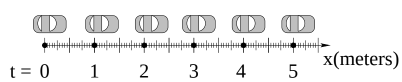


Figure 1.2: The location of the car on the new x-axis at each time are represented by dots.

We could then abstract this by measuring distances between snaps on the street and then fitting a scaled **coordinate line** on top of the path of the car as in figure 1.2:

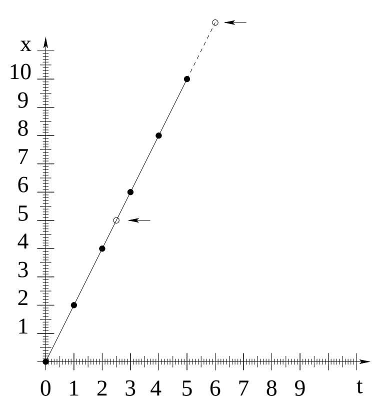


Figure 1.3: A graph of the position of the car as a function of time. The filled dot with an arrow represents a pretty darned good *guess* as to where the car was at time $t = 2.5$ seconds. The open dot with an arrow at $t = 6$ seconds is not as reliable. Why not?

Finally, we transform this into a graph of a **function** representing the position of the car $x(t)$ as in figure 1.3. Note how we can easily *extrapolate* the position of the car to a time outside of our series of observations, or *interpolate* its position in between our observations, subject to a few assumptions!

Believe it or not, we've invented one key component of physics! It's called ***kinematics***, and it is basically ***"geometric math with units"!*** We can abstract more information about the car from this graph. For example, the initial data – the speed of the car – is observable on the graph as:

$$v = \frac{\Delta x}{\Delta t} \tag{1.13}$$

which (having had a course of calculus) we recognize as being the *slope* of the straight line we drew interpolating the points.

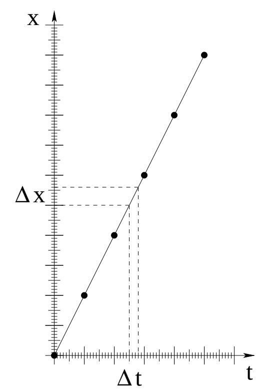


Figure 1.4: The speed of the car is the slope of its position as a function of time graph, $x(t)$.

This is illustrated (by indicating an interval in space Δx spanning two possible points on the curve and the corresponding interval in time Δt spanning) in figure 1.4.

Note that the car could be going in one of two directions along the x-axis: to the right (positive slope), or the left (negative slope). While we are watching the car going from left to right, our neighbor across the street is watching it go from right to left! Our graph alone is not useful, then, unless we somehow agree on a way of removing this ambiguity. We define the ***velocity*** of the car to be its speed ***plus the unique specification of the direction in which it moves!*** This can be simple/verbal – "to the right" (from a prespecified point of view) or complicated in multiple dimensions – "at 15° east of due north" or abstract – "in the positive x-direction" (on a specified coordinate frame).

"Speed" in physics is defined to be the *magnitude* of the velocity. Velocity is a *vector* quantity – one with magnitude *and* direction. Even in 1D motion, speed can go to the right or the left (+x or -x direction)!

Real cars, of course, move in ways that are more complicated than indicated in this simple example. In particular, they can speed up, slow down, and (eventually) change direction! In figure 1.5 I've drawn an $x(t)$ curve – also called the ***trajectory*** – of a hypothetical car as it *speeds up from rest* to achieve a constant speed. Note that we are now abstractly describing something that *never happened to any specific car* (but might) or that *generically* happens to *all cars* as they speed up or slow down! By building a *representation* of the world in our minds, we can start to reason *about* the world in consistent ways!

Considering the rate that the *velocity* changes lets us define a new quantity – the ***acceler-***

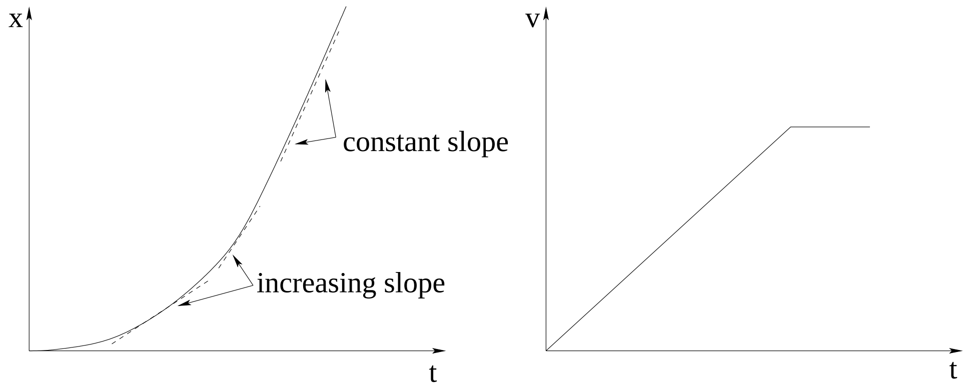


Figure 1.5: The graph of an *accelerating* car – one that is *speeding up* as it moves.

ation – that is the rate of change of the velocity with respect to time:

$$a = \frac{\Delta v}{\Delta t} \tag{1.14}$$

Acceleration is *also* a vector quantity. It can point in the same direction as the velocity (speeding it up), opposite to the direction (slowing it down), at right angles to the direction (changing its direction of motion) or some combination of these! In figure 1.5 the slope of the $x(t)$ trajectory is constantly *changing* as the car speeds up. This means that

$$v = \frac{\Delta x}{\Delta t} \tag{1.15}$$

is not a very good definition! It *changes* – quite possibly by a lot – as we choose different Δt intervals begun at the same point! How can we make our definition unique, and hence useful?

Newton solved this problem by *inventing calculus!* He defined the ***instantaneous velocity*** to be the infinitesimal increase in position divided by the infinitesimal increase in time over which it moves, or:

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt} \tag{1.16}$$

We're going to go right ahead and use this trick again, to also define the ***instantaneous acceleration***:

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} = \frac{d^2x}{dt^2} \tag{1.17}$$

In either case they are the *slopes* of the $x(t)$ or $v(t)$ curves, respectively.

We *gain a lot* by shrinking 50 meters or so of street to a few centimeters on a piece of paper and representing *real* motion of *real* cars on the *real* street with abstract symbols like x, t, some drawn coordinate axes, etc. On the street, it is too much, too big, to keep in our heads. On the paper, we can manipulate it and digest it and even invent imaginary cars.

Note that I omitted specifying the units *on* the graphs above. Lacking them, do we know whether t is measured in seconds, hours, or centuries? Is x in meters, centimeters, furlongs, rods? We have to provide a *legend* for this *map* (because that little picture *is* a kind of "dynamical map", one representing the motion in space and time, echoing and reinforcing a cognitive/conceptual map we are co-building in our *brains*). One key part of the legend is the *units* used for/in the quantities. To avoid a rather considerable ambiguity we *either* have to specify these units explicitly on the graph and in any mention of numerical quantities *or* use a convention.

1.3.2: Maps with Units

Let's spend a moment to generalize this to the point where we can talk about more than just cars on a one-dimensional street. Think about any *thing*, any entity that objectively exists in the real, visible, Universe. What defines the object and differentiates it from all of the *other* things that make up the Universe? Before we can talk about how the Universe and its contents change in time, we have to talk about how to describe its contents (and time itself) *at all*.

As I type this I'm looking over at just such a thing across the room from me, an object that I truly believe exists in the real Universe. To help you understand this object, I have to use *language*. I might tell you how large it is, what its weight is, what it looks like, where it is, how long it has been there, what it is for, and – of course – I have to use *words* to do this, not just nouns but a few adjectival modifiers, and speak of an "empty beer glass sitting on a table in my den just to my side", where now I have only to tell you just where my den is, where the table is in the den, and perhaps differentiate this particular beer glass from other beer glasses you might have seen. Eventually, if I use enough words, construct a sufficiently detailed map, make careful measurements, perhaps include a picture, I can convey to you a very precise *mental picture* of the beer glass, one sufficiently accurate that you can *visualize* just where, when and what it is (or was).

Of course this prose description is *not the glass itself!* If you like, the *map is not the territory*[46]! That is, it is an *informational representation* of the glass, a collection of symbols with an agreed upon meaning (where meaning in this context is a correspondence between the symbols and the presumed general sensory experience of the glass that one would have if one *looked* at the glass from my current point of view).

Constructing just such a map is the whole point of physics, only the map is not just of mundane objects such as a glass; it is the map of the whole world, the whole *Universe*. To the extent that this *worldview* is successful, especially in a predictive sense and not just hindsight, the physical map in your mind works well to predict the Universe you perceive through your sensory apparatus. A perfect understanding of physics (and a knowledge of certain data) is equivalent to a possessing a perfect map, one that precisely locates every thing within the Universe for all possible times.

Maps require several things. It is convenient, but not necessary, to have a set of single term descriptors, symbols for various "things" in the world the map is supposed to describe. So this symbol might stand for a house, that one for a bridge, still another one for a street or railroad crossing. Another *absolutely essential* part of a map is the actual *coordinates* of things that it is describing. The coordinate representation of objects drawn in on the map is supposed to exist in an accurate one-to-one but *abstract* correspondence with the *concrete* territory in the real world where the things represented by the symbols actually exist and move about[47].

Of course the symbols such as the term "beer glass" can *themselves* be abstractly modeled as points in some sort of space; Complex or composite objects with "simple" coordinates can

[46]This is an adage of a field called *General Semantics* – an approach to epistemology – and is something to keep carefully in mind while studying physics. Not even my direct perception of the glass is the glass itself; it is just a more complex and dynamical kind of map.

[47]Of course in the old days most actual maps were stationary, and one had to work hard to see "time" on them, but nowadays nearly everybody has or at least has seen GPS maps and video games, where things or the map coordinates themselves *move*.

be represented as a collection of far more coordinates for the smaller objects that make up the composite object. Ultimately one arrives at *elementary* objects, objects that are not (as far as we know or can tell *so far*) made up of other objects at all. The various kinds of elementary objects, the list of their variable properties, and their spatial and temporal coordinates are in some deep sense *all coordinates*, and every object in the universe can be thought of as a *point in or volume of this enormous and highly complex coordinate space!*

In this sense "coordinates" are the *fundamental adjectival modifiers* corresponding to the differentiating properties of "named things" (nouns) in the real Universe, where the term fundamental can also be thought of as meaning elementary or *irreducible* – adjectival properties that cannot be readily be expressed in terms of or derived from other adjectival properties of a given object/thing.

Physical coordinates are, then, basically mathematical numbers with units (and can be so considered even when they are discrete non-ordinal sets). At first we will omit most of the details from any given object we study to keep things simple. In fact, we will spend much of the first part of the course interested in only the three quantities included in table 1 below that set the scale for the coordinate system we need to describe the classical physics of a rather generic "location" in space (**where** the object is), a "time" (**when** it is there), and specify the "mass" (an intrinsic property representing **what** the object is, in the sense most important to dynamics).

You may well say to yourself "But what about the fact that it was a *beer glass* and not a *head of cabbage* or *car*? Doesn't that *matter*?" And of course, the answer is yes, but we won't be able to understand how to handle the differences until we've learned the physics of generic objects *without* all of these differences. Indeed, we'll eventually prove that all of the information we ignore is *not necessary* to predict at least part of the dynamical motion of any given object with great precision!

This is our first *idealization* – the treatment of an extended (composite) object as if it were itself an elementary object. This is called the *particle approximation*, and later we will *justify* this approximation *a posteriori* (after the fact) by showing that there really *is* a special point in a collective system of particles that behaves like a particle as far as Newton's Laws are concerned. For the time being, then, objects such as porpoises, puppies, and ponies are all idealized and will be treated as *particles*[48]. We'll talk more about particles in a page or two.

This is why we need *units* to describe intervals or values in all three coordinates so that we can talk or think about those particles (idealized objects) in ways that don't depend on the listener. The default units used in this course – ***by convention*** – come from the **SI Coordinate System** (Système International (d'unités)). Each **dimension** of a physics problem – quantities like mass, length, time – get their own unique set of **default symbols** and their very own **default units**:

All other units for at least a short while will be expressed/expressible in terms of these three.

[48]I teach physics in the summers at the Duke Marine Lab, where there are porpoises and wild ponies visible from the windows of our classroom. Puppies I threw in for free because they are cute and also begin with "p". However, you can think of a particle as a baseball or bullet or ball bearing if you prefer less cute things that begin with the letter "b", which is a reflection transformed "p".

Dimension	Symbols	Units
Length	x,y,z,r,ℓ	meters
Time	t,T,τ	seconds
Mass	m,M	kilograms

Table 1: The SI units used by default in the first semester textbook of this course. An additional unit is needed in the second semester textbook!

For example units of *velocity* will be meters per second, units of *force* will be kilogram-meters per second squared. We will often give names to some of these combinations, such as the SI units of force:

$$1 \text{ Newton} = \frac{\text{kg-m}}{\text{sec}^2} \tag{1.18}$$

Later you may learn of other irreducible coordinates that describe elementary particles or extended macroscopic objects in physics such as (next semester) electrical charge, as well as additional quantities such as energy, momentum, angular momentum, electrical current, and more, that can be expressed in terms of these basic coordinates.

As for what the quantities that these units represent *are* – well, that's a tough question to answer. I know how to *measure* distances between points in space and times between events that occur in space, using things like meter sticks and stopwatches, but as to just what the space and time in which these events are embedded *really are* I'm as clueless as a caveman. It's probably best to just define distance as that which we might measure with a meter stick or other "standard" of length, time as that which we might measure with a clock or other "standard" of time, and mass that which we might measure compared to some "standard" of mass using methods we'll have to figure out below. Existential properties cannot really be defined, only observed, quantified, and understood in the *context* of a complete, consistent system, the physical worldview, the map we construct that *works* to establish a useful semantic representation of that which we observe.

Sorry if that's difficult to grasp, but there it is. It's just as difficult for me (after most of a lifetime studying physics) as it is for you right now. Dictionaries are, after all, written in words that are *in* the dictionaries and hence are self-referential and in some deep sense should be abstract, arbitrary, meaningless – yet somehow we learn to speak and understand them. So it is with language, so it will be for you with the "language" of physics, and the process in both cases takes time, experience, and *practice!*

Coordinates are enormously powerful ideas, the very essence of mapmaking and knowledge itself. To assist us in working with them, we introduce the notion of *coordinate frame* – a *system* of all of the relevant coordinates that describe at least the position of a particle (in one, two or three dimensions, usually). In figure 1.6 is a picture of a simple single particle with mass m (that might represent my beer glass sliding around on a table or an ant walking along on a floor)) on a set of *coordinates* that describes at least part of the actual space where my (say) beer glass is sitting. The solid line on this figure represents the *trajectory* of my glass as it moves about.

Armed with a watch, an apparatus for measuring mass, meter sticks and some imagination, one can imagine, say, a virtual car driving around town along its virtual trajectory and compare

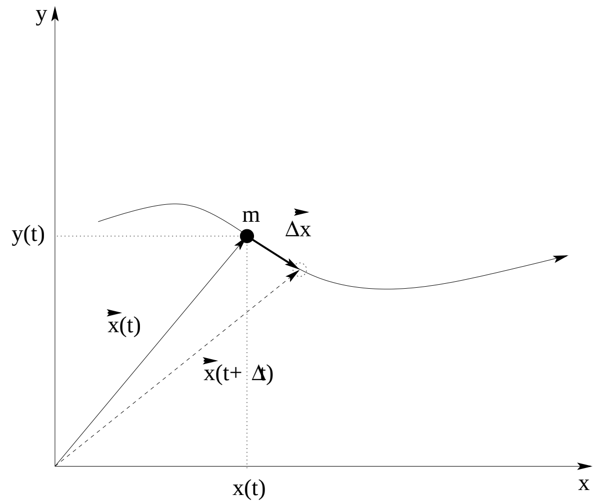


Figure 1.6: Coordinatized visualization of the motion of a particle of mass m along a trajectory $\vec{x}(t)$. Note that in a short time Δt the particle's position changes from $\vec{x}(t)$ to $\vec{x}(t+\Delta t)$.

its motion in our *conceptual map*[49] with the correspondent happenings in the world outside of our minds where the real car moves along a real track.

Note well that we *idealize* things like a car by treating the whole thing as a *single object* with a *single position* (located somewhere "in the middle") when we know that it is *really* made up of steering wheels and bucket seats that are "objects" in their own right that are further assembled into a "car" All of these wheels and panels, nuts and bolts are made up of still smaller objects – molecules – and molecules are made up of atoms, and atoms are made of protons and neutrons and electrons, and protons and neutrons are made up of quarks, and we don't really *know* for certain if electrons and quarks are truly elementary particles or are themselves composite objects[50].

Later in this semester we will *formally justify* our ability to do this, and we will improve on our description of things like cars and wheels and solid composite objects by learning how they can move, rotate, and even explode into little bits of car and still have *some* parts of their collective coordinate motion that behaves as though the ex-car is still a "single point-like object".

In the meantime, we will simply begin with this idealization and treat discrete solid objects as *particles* – masses that are at a single point in space at a single time. So we will treat objects such as planets, porpoises, puppies, people, baseballs and blocks, cars and cannonballs and much more as if they have a single mass and a single spatial location at any single instant in time – as a particle. One advantage of this is that the mathematical abstraction for the location and mass of these "particles" become *functions* of time[51] and we establish a way of adding *other* descriptors in a consistent way to our abstraction as we discover them.

[49]This map need not be paper, in other words – I can sit here and visualize the entire drive from my house to the grocery store, over time. Pictures of trajectories on paper are just ways we help our brains manage this sort of understanding.

[50]Although the currently accepted belief is that they are. However, it would take only one good, reproducible experiment to make this belief less plausible, more probably false. Evidence talks, belief walks.

[51]Recall that a function is a quantity that depends on a set of argument(s) that is single-valued, that is, has a single value for each unique value of its argument(s).

1.3.3: Coordinates and Vector Quantities

In physical dynamics we will initially be concerned with finding the *trajectories* of particles (in systems of particles that initially will be a single particle) – the position of each particle in the system expressed as a *function* of time. We can write the trajectory as a *vector function* on a spatial coordinate system (e.g. cartesian coordinates in 2 dimensions):

$$\vec{\boldsymbol{x}}(t) = x(t)\hat{\boldsymbol{x}} + y(t)\hat{\boldsymbol{y}} \tag{1.19}$$

Note that $\vec{\boldsymbol{x}}(t)$ stands for a vector from the origin to the particle, where $x(t)$ by itself (no boldface or vector sign) stands for the x-component of this vector. An example trajectory is *visualized* in figure 1.6 (where it might stand for the trajectory of my car, or beer glass, treated as a particle). In all of the problems we work on throughout the semester, *visualization* will be a key component of the solution.

The human brain doesn't, actually, excel at being able to keep all of these details onboard in your "mind's eye", the virtual visual field of your imagination. Consequently, you must *always* draw figures, usually with coordinates, forces, and other "decorations", when you solve a physics problem. The paper (or blackboard, or whiteboard, or computer screen) then becomes an extension of your brain – a kind of "scratch space" that augments your visualization capabilities and your sequential symbolic reasoning capabilities. To put it bluntly, you are *more intelligent* when you reason with a piece of paper and pen than you are when you are forced to rely on your brain alone. To succeed in physics, you need all of the intelligence you can get, and you need to *synthesize* solutions using both halves of your brain: visualization and sequential reasoning. Paper and pen facilitate this process and one of the most important lessons of this course is how to *use* them to attain the full benefit of the added intelligence they enable not just in physics problems, but everywhere else as well.

As we previously discussed, if we know the trajectory function of a particle, we know a lot of other things too. Since we know *where* it is at any given time, we can easily tell *how fast it is going* in *what direction*. This combination of the speed of the particle and its direction forms a vector called the *velocity* of the particle. Speed, we all hopefully know from the previous section and from our experience in real life doing things like driving cars, is a measure of how far you go in a certain amount of time, expressed in units of *distance* (length) divided by *time*. Miles per hour. Furlongs per fortnight. Or, in the SI units we will use in this physics course, *meters per second*[52].

When moving around in two or three dimensions in space described by some coordinate frame, the *average* velocity of the particle will now by definition be the *vector* change in its position $\Delta\vec{\boldsymbol{x}}$ in some time Δt divided by that time:

$$\vec{\boldsymbol{v}}_{\rm avg} = \frac{\Delta\vec{\boldsymbol{x}}}{\Delta t} \tag{1.20}$$

Sometimes average velocity is useful, but often, even usually, it is not. It can be a rather poor measure for how fast a particle is actually *moving* at any given time, especially if averaged over times that are long enough for interesting *changes* in the motion to occur.

[52]A good rule of thumb for people who have a practical experience of speeds in miles per hour trying to visualize meters per second is that ***1 meter per second*** is approximately equal to ***9/4 miles per hour***, hence four meters per second is nine miles per hour. A cruder but still quite useful approximation is meters per second equals miles per hour/2, or mph = mps times 2.

For example, I might get in my car and drive around a racetrack at speed of 50 meters per second – really booking it, tires squealing on the turns, smoke coming out of my engine (at least if I tried this in *my* car, as it would likely explode if I tried to go 112 mph for any extended time), and screech to a halt right back where I began. My *average* velocity is then *zero* – I'm back where I started! That zero is a poor explanation for the heat waves pulsing out from under the hood of the car and the wear on its tires.

More often we will be interested in the *instantaneous velocity* of the particle. This is basically the average velocity, averaged over as small a time interval as possible – one so short that it is just long enough for the car to move at all. Calculus permits us to take this limit, and indeed uses just this limit as the *definition* of the *derivative*. We thus define the instantaneous velocity vector as the time-derivative of the position vector:

$$\vec{\boldsymbol{v}}(t) = \lim_{\Delta t \to 0} \frac{\vec{\boldsymbol{x}}(t+\Delta t) - \vec{\boldsymbol{x}}(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta \vec{\boldsymbol{x}}}{\Delta t} = \frac{d\vec{\boldsymbol{x}}}{dt} \tag{1.21}$$

Sometimes we will care about "how fast" a car is going but not so much about the direction. **Speed** is defined to be the *magnitude* of the velocity vector:

$$v(t) = |\vec{\boldsymbol{v}}(t)| \tag{1.22}$$

We could say more about it, but I'm guessing that you already have a pretty good intuitive feel for speed if you drive a car and pay attention to how your speedometer reading corresponds to the way things zip by or crawl by outside of your window.

The reason that average velocity is a poor measure is that (of course) our cars speed up and slow down and change direction, often. Otherwise they tend to run into things, because it is usually not possible to travel in perfectly straight lines at only one speed and drive to the grocery store. To see how the *velocity* changes in time, we will need to consider the *acceleration* of a particle, or the rate at which the *velocity* changes. As before, we can easily define an average acceleration over a possibly long time interval Δt as:

$$\vec{\boldsymbol{a}}_{\text{avg}} = \frac{\vec{\boldsymbol{v}}(t+\Delta t) - \vec{\boldsymbol{v}}(t)}{\Delta t} = \frac{\Delta \vec{\boldsymbol{v}}}{\Delta t} \tag{1.23}$$

Also as before, this average is usually a poor measure of the actual acceleration a particle (or car) experiences. If I am on a straight track at rest and stamp on the accelerator, burning rubber until I reach 50 meters per second (112 miles per hour) and then stamp on the brakes to quickly skid to a halt, tires smoking and leaving black streaks on the pavement, my *average* acceleration is once again zero, but there is only one brief interval (between taking my foot off of the accelerator and before I pushed it down on the brake pedal) during the trip where my *actual* acceleration was anything *close* to zero. Yet, my average acceleration is zero.

Things are just as bad if I go around a circular track at a constant speed! As we will shortly see, in that case I am *always* accelerating towards the center of the circle, but my average acceleration is systematically approaching zero as I go around the track more and more times.

From this we conclude that the acceleration that really matters is (again) the limit of the average over very *short* times; the time derivative of the velocity. This limit is thus defined to be the *instantaneous* accleration:

$$\vec{\boldsymbol{a}}(t) = \lim_{\Delta t \to 0} \frac{\Delta \vec{\boldsymbol{v}}}{\Delta t} = \frac{d\vec{\boldsymbol{v}}}{dt} = \frac{d^2\vec{\boldsymbol{x}}}{dt^2}, \tag{1.24}$$

the acceleration of the particle *right now*.

Obviously we could continue this process and define the time derivative of the acceleration – the third order time derivative of the trajectory[53] – and even higher order derivatives, but *as it turns out*, the dynamic principle that appears sufficient to describe pretty much all classical motion will involve *force* and *acceleration* only. Pretty much all of the math we do (at least at first) will involve solving *backwards* from a knowledge of the acceleration to a knowledge of the velocity and position vectors via *integration* or more generally (later) *solving the second-order differential equation of motion* with no need to consider third or higher order derivatives at all.

We are now prepared to formulate this dynamical principle – Newton's Second Law. While we're at it, let's study his First and Third Laws too – might as well collect the complete set...

1.4: Newton's Laws

The following are Newton's Laws as you will need to know them to both solve problems and answer conceptual questions in this course. Recall from our former discussion that they are ***postulates*** and hence incapable of being derived per se except from other, equivalent postulates, and are justified in science only by their success at describing the real world. Note well that they are framed in terms of the *spatial coordinates* defined in the previous section plus mass and time.

a) **Law of Inertia:** Objects at rest or in uniform motion (at a constant velocity) in an *inertial reference frame* remain so unless acted upon by an unbalanced (net, total) force. We can write this algebraically as:

$$\vec{\boldsymbol{F}} = \sum_i \vec{\boldsymbol{F}}_i = 0 = m\vec{\boldsymbol{a}} = m\frac{d\vec{\boldsymbol{v}}}{dt} \quad \Rightarrow \quad \vec{\boldsymbol{v}} = \text{constant vector} \tag{1.25}$$

b) **Law of Dynamics:** The total force applied to an object is directly proportional to its acceleration in an *inertial reference frame*. The constant of proportionality is called the **mass** of the object. We write this algebraically as:

$$\vec{\boldsymbol{F}} = \sum_i \vec{\boldsymbol{F}}_i = m\vec{\boldsymbol{a}} = \frac{d(m\vec{\boldsymbol{v}})}{dt} = \frac{d\vec{\boldsymbol{p}}}{dt} \tag{1.26}$$

where we introduce the *momentum* of a particle, $\vec{\boldsymbol{p}} = m\vec{\boldsymbol{v}}$, in the last way of writing it although we won't study this form until chapter/week 4.

c) **Law of Reaction:** If object B exerts a (named) force $\vec{\boldsymbol{F}}_{AB}$ on object B *along a line connecting the two objects*, then object A exerts an equal and opposite reaction force of $\vec{\boldsymbol{F}}_{BA} = -\vec{\boldsymbol{F}}_{AB}$ on object B. Because we can have a lot more than just two particles A and B, we write this algebraically as:

$$\vec{\boldsymbol{F}}_{ij} = -\vec{\boldsymbol{F}}_{ji} \tag{1.27}$$

$$\text{(or)} \quad \sum_{i,j} \vec{\boldsymbol{F}}_{ij} = 0 \tag{1.28}$$

[53] A quantity that actually does have a name – it is called the *jerk* – but we won't need it.

where i and j are arbitrary particle labels. The latter form will be useful to us later; it means that the sum of all *internal* forces between particles in any closed system of particles cancels!.

Note that these laws are *not all independent* as mathematics goes. The first law is a clear and direct consequence of the second. The third is not – it is an independent statement. The first law *historically*, however, had an important purpose. It *rejected* the dynamics of Aristotle, introducing the new idea of *intertia* where an object in motion continues in that motion unless acted upon by some external agency. This is directly opposed to the Aristotelian view that things only moved when acted on by an external agency and that they naturally came to *rest* when that agency was removed. A second important purpose of the first law is that – together with the third law – it helps us *define* an ***inertial reference frame*** as a frame where the first law is true.

The second law is our basic dynamical principle. It tells us how to go from a problem description (in words) plus a knowledge of the force laws of nature to an "equation of motion" (typically a statement of Newton's second law). The equation of motion, generally solved for the acceleration, becomes a *kinematical* equation from which we can develop a full knowledge of the motion using mathematics guided by our experience and physical intuition.

The third law leads (as we shall see) to the surprising result that *systems of particles behave collectively like a particle!* This is indeed fortunate! We know that something like a baseball is really made up of a *lot* of teensy particles itself, and yet it obeys Newton's Second law as if it *is* a particle. We will use the third law to derive this and the closely related *Law of Conservation of Momentum* in a later week of the course.

An inertial reference frame is a coordinate system (or frame) that is either at rest or moving at a constant speed, a *non-accelerating* frame of reference. For example, the ground, or lab frame, is a coordinate system at rest relative to the approximately non-accelerating ground or lab and is considered to be an inertial frame to a good approximation. A (coordinate system inside a) car travelling at a constant speed *relative* to the ground, a spaceship coasting in a region free from fields, a railroad car rolling on straight tracks at constant speed are also inertial frames. A (coordinate system inside a) car that is accelerating (say by going around a curve), a spaceship that is accelerating, a freight car that is speeding up or slowing down – these are all examples of *non*-inertial frames. All of Newton's laws suppose an inertial reference frame (yes, the third law too) and are generally *false* for accelerations evaluated in an *accelerating* frame as we will prove and discuss next week.

In the meantime, please be sure to learn the statements of the laws *including* the condition "in an inertial reference frame", in spite of the fact that you don't yet really understand what this means and why we include it. Eventually, it will be the *other* important use of Newton's First and Third Laws – to *define* an inertial reference frame as any frame where an object remains in a state of uniform motion if no named forces arising from pairwise interactions act on it!

You'll be glad that you did.

1.5: Forces

Classical dynamics at this level, in a nutshell, is very simple. Find the total force on an object. Use Newton's second law to obtain its acceleration (as a differential equation of motion). Solve the equation of motion by direct integration or otherwise for the position and velocity.

That's it!

Well, except for answering those pesky *questions* that we typically ask in a physics problem, but we'll get to that later. For the moment, the next most important problem is: how do we evaluate the total force?

To answer it, we need a knowledge of the forces at our disposal, the force laws and rules that we are likely to encounter in our everyday experience of the world. Some of these forces are *fundamental* forces – *elementary* forces that we call "laws of nature" because the forces themselves aren't caused by some *other* force, they are themselves the actual causes of dynamical action in the visible Universe. Other force laws aren't quite so fundamental – they are more like "approximate rules" and aren't exactly correct. They are also usually derivable from (or at least understandable from) the elementary natural laws, although it may be quite a lot of work to do so.

We quickly review the forces we will be working with in the first part of the course, both the forces of nature and the force *rules* that apply to our everyday existence in approximate form.

1.5.1: The Forces of Nature

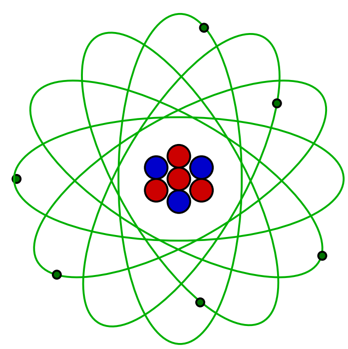


Figure 1.7: A very simple "toy model" of an atom as a nucleus with electrostatically bound (green) orbiting electrons. The nucleus itself is made up of protons (blue) and neutrons (red) "glued" together with the strong nuclear force.

At this point in your life, you almost certainly know that all normal matter of your everyday experience is made up of *atoms*. Most of you also know that an atom itself is made up of a positively charged *atomic nucleus* that is very tiny indeed surrounded by a cloud of negatively charged *electrons* that are much lighter. Atoms in turn bond together to make molecules, atoms or molecules in turn bind together (or not) to form gases, liquids, solids – material "things", including those macroscopic solid things like puppies that we are so far planning to *treat* as particles.

The *actual* elementary particles from which they are made are much tinier than atoms. It

Particle	Location	Size
Up or Down Quark	Nucleon (Proton or Neutron)	pointlike
Proton	Nucleus	10^{-15} meters
Neutron	Nucleus	10^{-15} meters
Nucleus	Atom	10^{-15} meters
Electron	Atom	pointlike
Atom	Molecules or Objects	$\sim 10^{-10}$ meters
Molecule	Objects	$> 10^{-10}$ meters

Table 2: Basic massive building blocks of normal matter as of 2021, subject to change as we discover and understand more about the Universe, ignoring pesky things like neutrinos, photons, gluons, heavy vector bosons, and heavier leptons in the "Standard Model" that physics majors (at least) will have to learn about later...

is worth providing a *greatly simplified table* of the "stuff" from which normal atoms (and hence molecules, and hence we ourselves) are made.

In table 2, up and down quarks and electrons are so-called *elementary particles* – things that are not made up of something else but are fundamental components of nature. Quarks bond together three at a time to form *nucleons* as shown in figure 1.8 – a proton is made up of "up-up-down" quarks and has a charge of $+e$, where e is the elementary electric charge. A neutron is made up of "up-down-down" and has no charge.

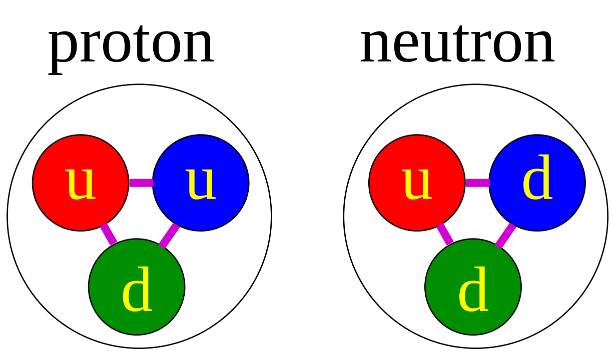


Figure 1.8: A very simple "toy model" of a proton or neutron made up of three quarks.

Neutrons and protons, in turn, bond together to make an atomic nucleus. The simplest atomic nucleus is the hydrogen nucleus, which can be as small as a single proton, or can be a proton bound to one neutron (deuterium) or two neutrons (tritium). No matter how many protons and neutrons are bound together, the atomic nucleus is *small* – order of 10^{-15} meters in diameter[54]. The quarks, protons and neutrons are bound together by means of a force of nature called the *strong nuclear interaction*, which is the strongest force we know of relative to the mass of the interacting particles.

The positive nucleus combines with electrons (which are negatively charged and around 2000 times lighter than a proton) to form an *atom*. The force responsible for this binding is the *electromagnetic* force, another force of nature (although in truth nearly all of the interaction is

[54]...with the possible exception of neutrons bound together by *gravity* to form *neutron stars*. Those can be thought of, very crudely, as very large nuclei.

electrostatic in nature, just one part of the overall electromagnetic interaction). The form of the electrostatic (Coulomb) force between the nucleus and the electrons and electrons with each other is:

$$\vec{\boldsymbol{F}}_{12} = \frac{k_e q_1 q_2}{r_{12}^2}\hat{\boldsymbol{r}}_{12} \tag{1.29}$$

where k_e is a constant of nature that sets the scale of the interaction in terms of the units used in the variables in the equation, $\vec{\boldsymbol{F}}_{12}$ should be read as "the force on object 1 due to object 2", q_1, q_2 are the charges of the two interacting objects and $\vec{\boldsymbol{r}}_{12}$ is the relative vector *from* object 2 *to* object 1. We will not spend any great amount of time studying this force *this* semester/course as the entire *second* semester/course of introductory physics is typically devoted to the study of electricity and magnetism and the related topic of optics.

The light electrons repel one another electrostatically almost as strongly as they are attracted to the nucleus that anchors them. They also obey the *Pauli exclusion principle* which causes them to avoid one another's company. These things together cause atoms to be much larger than a nucleus, and to have interesting "structure" that gives rise to chemistry and molecular bonding and (eventually) life.

Inside the nucleus (and its nucleons) there is another force that acts at very short range. This force can cause e.g. neutrons to give off an electron and turn into a proton or other strange things like that. This kind of event changes the atomic number of the atom in question and is usually accompanied by *nuclear radiation*. We call this force the *weak nuclear force*. The two nuclear forces both exist only at *very short length scales*, basically in the quantum regime inside an atomic nucleus, where we cannot easily see them using the kinds of things we'll talk about this semester. For our purposes it is enough that they exist and bind stable nuclei together so that those nuclei in turn can form atoms, molecules, objects, *us*.

Our picture of normal matter, then, is that it is made up of *atoms* that may or may not be bonded together into molecules, with three forces all significantly contributing to what goes on inside a nucleus and only one that is predominantly relevant to the electronic structure of the atoms themselves.

There is, however, a fourth force (that we know of – there may be more still, but four is all that we've been able to positively identify and understand). That force is **gravity**. Gravity is a bit "odd". It is a very long range, but *very weak* force – by far the weakest force of the four forces of nature. It only is signficant when one builds *planet* or *star* sized objects, where it can do anything from simply binding an atmosphere to a planet and causing moons and satellites to go around it in nice orbits to bringing about the *catastrophic collapse of a dying star!* The physical law for gravitation will be studied over an entire week of work – *later* in the course. I put it down now just for completeness, but initially we'll focus on the force *rules* in the following section.

$$\vec{\boldsymbol{F}}_{12} = -\frac{G m_1 m_2}{r_{12}^2}\hat{\boldsymbol{r}}_{12} \tag{1.30}$$

Note that this is very similar to the Coulomb force above; G is a constant of nature, the m's are the *masses* of the interacting objects, etc. Again, don't worry too much about what all of these symbols mean and what the value of G is – we'll get to all of that but not now.

Since we *live* on the surface of a planet, to *us* gravity will be an important force, but the forces we experience every day and we ourselves are primarily electromagnetic phenomena, with a bit of help from quantum mechanics to give all that electromagnetic stuff just the right structure.

Let's summarize this in a short table of forces of nature, strongest to weakest:

a) Strong Nuclear

b) Electromagnetic

c) Weak Nuclear

d) Gravity

Note well: It is possible that there are more forces of nature waiting to be discovered. Because physics is not a *dogma*, this presents no real problem. If a new force of nature (or radically different way to view the ones we've got) emerges as being consistent with observation and predictive, and hence possibly/plausibly true and correct, we'll simply give the discoverer a Nobel Prize, add their name to the "pantheon of great physicists", add the force itself to the list above, and move on. Science, as noted above, is a *self-correcting* system of reasoning, at least when it is done right.

1.5.2: Force Rules

The following set of force rules will be used both in this chapter and throughout this course. All of these rules can be derived or understood (with some effort) from the forces of nature, that is to say from "elementary" natural laws, but are not *quite* laws themselves.

a) **Gravity** (near the surface of the earth):

$$F_g = mg \tag{1.31}$$

The direction of this force is ***down***, so one could write this in vector form as $\vec{\boldsymbol{F}}_g = -mg\hat{\boldsymbol{y}}$ in a coordinate system such that up is the $+y$ direction. This rule follows from Newton's Law of Gravitation, the elementary law of nature in the list above, evaluated "near" the surface of the earth where it is varies only very slowly with height above the surface (and hence is "constant") as long as that height is small compared to the radius of the Earth.

The *measured* value of g (the gravitational "constant" or gravitational field close to the Earth's surface) thus *isn't* really constant – it actually varies weakly with latitude and height and the local density of the earth immediately under your feet and is pretty complicated[55] . Some "constant", eh?

Most physics books (and the wikipedia page I just linked) give g's value as something like:

$$g \approx 9.81 \quad \frac{\text{meters}}{\text{second}^2} \tag{1.32}$$

[55]Wikipedia: http://www.wikipedia.org/wiki/Gravity_of_Earth. There is a very cool "rotating earth" graphic on this page that shows the field variation in a color map. This page goes into much more detail than I will about the causes of variation of "apparent gravity".

(which is sort of an average of the variation) but in this class to the extent that we do arithmetic with it we'll just use

$$g \approx 10 \quad \frac{\text{meters}}{\text{second}^2} \tag{1.33}$$

because hey, so it makes a 2% error. That's not very big, really – you will be lucky to measure g in your labs to within 2%, and it is *so* much easier to multiply or divide by 10 than 9.80665.

b) **The Spring** (Hooke's Law) in one dimension:

$$F_x = -k\Delta x \tag{1.34}$$

This force is directed back to the equilibrium point (the end of the unstretched spring where the mass is attached) in the *opposite direction* to Δx, the displacement of the mass on the spring away from this equilibrium position. This rule arises from the primarily electrostatic forces holding the atoms or molecules of the spring material together, which tend to linearly oppose *small* forces that pull them apart or push them together (for reasons we will understand in some detail later).

c) The **Normal Force**:

$$F_\perp = N \tag{1.35}$$

This points perpendicular and away from solid surface, magnitude sufficient to oppose the force of contact *whatever it might be!* This is an example of a *force of constraint* – a force whose magnitude is determined by the *constraint* that one solid object cannot generally interpenetrate another solid object, so that the solid surfaces exert whatever force is needed to prevent it (up to the point where the "solid" property itself fails). The physical basis is once again the electrostatic molecular forces holding the solid object together, and *microscopically* the surface deforms, however slightly, more or less like a spring to create the force.

d) **Tension** in an Acme (massless, unstretchable, unbreakable) string:

$$F_s = T \tag{1.36}$$

This force simply transmits an *attractive* force between two objects on opposite ends of the string, in the directions of the taut string at the points of contact. It is another constraint force with no fixed value. Physically, the string is like a spring once again – it microscopically is made of bound atoms or molecules that pull ever so slightly apart when the string is stretched until the restoring force balances the applied force.

e) **Static Friction**

$$f_s < \mu_s N \tag{1.37}$$

(directed opposite to the *relative* net shear force parallel to the surfaces in contact). This is another *variable* force of constraint, as large as it needs to be to keep the object in question travelling at the same speed as the surface it is in contact with, up to the *maximum* value static friction can exert before the object starts to slide where we will assert that it slides when the shear force equals or exceeds $\mu_s N$. This force arises from mechanical interlocking at the microscopic level plus the electrostatic molecular forces that hold the surfaces themselves together.

f) **Kinetic Friction**

$$f_k = \mu_k N \tag{1.38}$$

(opposite to direction of relative sliding motion of surfaces and parallel to surface of contact). This force *does* have a fixed value when the right conditions (sliding) hold. This force arises from the forming and breaking of microscopic adhesive bonds between atoms on the surfaces plus some mechanical linkage between the small irregularities on the surfaces.

g) **Fluid Forces, Pressure**: A fluid in contact with a solid surface (or anything else) in general exerts a force on that surface that is related to the *pressure* of the fluid:

$$F_P = PA \tag{1.39}$$

which you should read as "the force exerted by the fluid on the surface is the pressure in the fluid times the area of the surface". If the pressure varies or the surface is curved one may have to use calculus to add up a total force. In general the direction of the force exerted is *perpendicular* to the surface. An object at rest in a fluid often has balanced forces due to pressure. The force arises from the molecules in the fluid literally bouncing off of the surface of the object, transferring momentum (and exerting an average force) as they do so. We will study this in some detail and will even derive a kinetic model for a gas that is in good agreement with real gases.

h) **Drag Forces**:

$$F_d = -bv^n \tag{1.40}$$

(directed opposite to relative velocity of motion through fluid, n usually between 1 (low velocity) and 2 (high velocity). It arises in part because the surface of an object moving through a fluid is literally bouncing fluid particles off in the leading direction while moving away from particles in the trailing direction, so that there is a differential pressure on the two surfaces, in part from "kinetic friction" that exerts a force component parallel to a surface in relative motion to the fluid. It is really pretty complicated – so complicated that we can only write down a specific, computable expression for it for very simple geometries and situations. Still, it is a very important and ubiquitous force and we'll try to gain some appreciation for it along the way.

1.6: Force Balance – Static Equilibrium

Before we start using dynamics at all, let us consider what happens when all of the forces acting on an object *balance*. That is, there are several non-zero (vector) forces acting on an object, but those forces sum up to zero force. In this case, Newton's *First* Law becomes very useful. It tells us that the object in question will remain at rest if it is initially at rest. We call this situation where the forces are all balanced *static force equilibrium*:

$$\vec{F}_{\rm tot} = \sum_i \vec{F}_i = m\vec{a} = 0 \tag{1.41}$$

This works both ways; if an object is at rest and stays that way, we can be certain that the forces acting on it balance!

We will spend some time later studying static equilibrium in general once we have learned about both forces and torques, but for the moment we will just consider a single example of what is after all a pretty simple idea. This will also serve as a short introduction to one of the forces listed above, *Hooke's Law* for the force exerted by a spring on an attached mass.

Example 1.6.1: Spring and Mass in Static Force Equilibrium

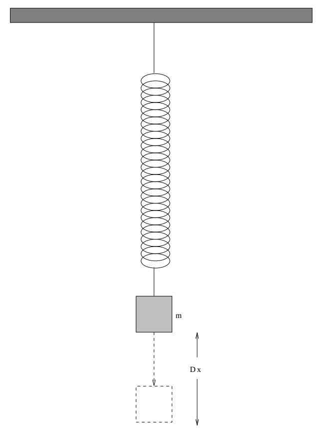


Figure 1.9: A mass m hangs on a spring with spring constant k. We would like to compute the amount Δx by which the string is stretched when the mass is at rest in static force equilibrium.

Suppose we have a mass m hanging on a spring with spring constant k such that the spring is stretched out some distance Δx from its unstretched length. This situation is pictured in figure 1.9.

We will learn how to really *solve* this as a dynamics problem later – indeed, we'll spend an entire week on it! Right now we will just write down Newton's laws for this problem so we can find a. Let the x direction be *up*. Then (using Hooke's Law from the list above):

$$\sum F_x = -k(x - x_0) - mg = ma_x \tag{1.42}$$

or (with $\Delta x = x - x_0$, so that Δx is negative as shown)

$$a_x = -\frac{k}{m}\Delta x - g \tag{1.43}$$

Note that this result doesn't depend on where the origin of the x-axis is, because x and x_0 both change by the same amount as we move it around. In most cases, we will find the equilibrium position of a mass on a spring to be the most convenient place to put the origin, because then x and Δx are the same!

In static equilibrium, $a_x = 0$ (and hence, $F_x = 0$) and we can solve for Δx:

$$\begin{aligned} a_x = -\frac{k}{m}\Delta x - g &= 0 \\ \frac{k}{m}\Delta x &= g \\ \Delta x &= \frac{mg}{k} \end{aligned} \tag{1.44}$$

You will see this result appear in several problems and examples later on, so bear it in mind.

1.7: Simple Motion in One Dimension

Finally! All of that preliminary stuff is done with. If you actually *read and studied* the chapter up to this point (many of you will not have done so, and you'll be SORRReeee...) you should:

a) Know Newton's Laws well enough to recite them on a quiz – yes, I usually just put a question like "What are Newton's Laws" on quizzes just to see who can recite them perfectly, a really useful thing to be able to do given that we're going to use them hundreds of times in the next 12 weeks of class, next semester, and beyond; and

b) Have at least *started* to commit the various force rules we'll use this semester to memory.

I don't generally encourage rote memorization in this class, but for a *few* things, usually very fundamental things, it can help. So if you haven't done this, go spend a few minutes working on this before starting the next section.

All done? Well all *rightie* then, let's see if we can actually *use* Newton's Laws (usually Newton's *Second* Law, our dynamical principle) and force rules to solve *problems*. We will start out very gently, trying to understand motion in one dimension (where we will not at first need multiple coordinate dimensions or systems or trig or much of the other stuff that will complicate life later) and then, well, we'll complicate life later and try to understand what happens in 2+ dimensions.

Here's the basic structure of a physics problem. You are given a physical description of the problem. A mass m at rest is dropped from a height H above the ground at time $t = 0$; what happens to the mass as a function of time? From this description you must visualize what's going on (sometimes but not always aided by a figure that has been drawn for you representing it in some way). You must select a coordinate system to use to describe what happens. You must write Newton's Second Law in the coordinate system for all masses, being sure to include *all* forces or force rules that contribute to its motion. You must solve Newton's Second Law to find the accelerations of all the masses (equations called the *equations of motion* of the system). You must solve the equations of motion to find the *trajectories* of the masses, their positions as a function of time, as well as their velocities as a function of time if desired. Finally, armed with these trajectories, you must *answer all the questions* the problem poses using algebra and reason and – rarely in this class – arithmetic!

Simple enough.

Let's put this simple solution methodology to the test by solving the following one dimensional, single mass example problem, and then see what we've learned.

Example 1.7.1: A Mass Falling from Height H

Let's solve the problem we posed above, and as we do so develop a *solution rubric* – a *recipe* for solving *all problems involving dynamics*[56]*!* The problem, recall, was to drop a mass m from rest from a height H, algebraically find the *trajectory* (the position function that solves the equations of motion) and *velocity* (the time derivative of the trajectory), and then answer any questions that might be asked using a mix of **algebra, intuition, experience and common sense**. For this first problem we'll postpone actually asking any question until we have these solutions so that we can see what *kinds* of questions one might reasonably ask and be able to answer.

The first step in solving *this or any physics problem* is to *visualize what's going on!* Mass m? Height H? Drop? Start at rest? Fall? All of these things are *input data* that *mean something* when translated into algebraic "physicsese", the language of physics, but in the end we have to *coordinatize* the problem (choose a coordinate system in which to do the algebra and solve our equations for an answer) and to choose a good one we need to *draw a representation* of the problem.

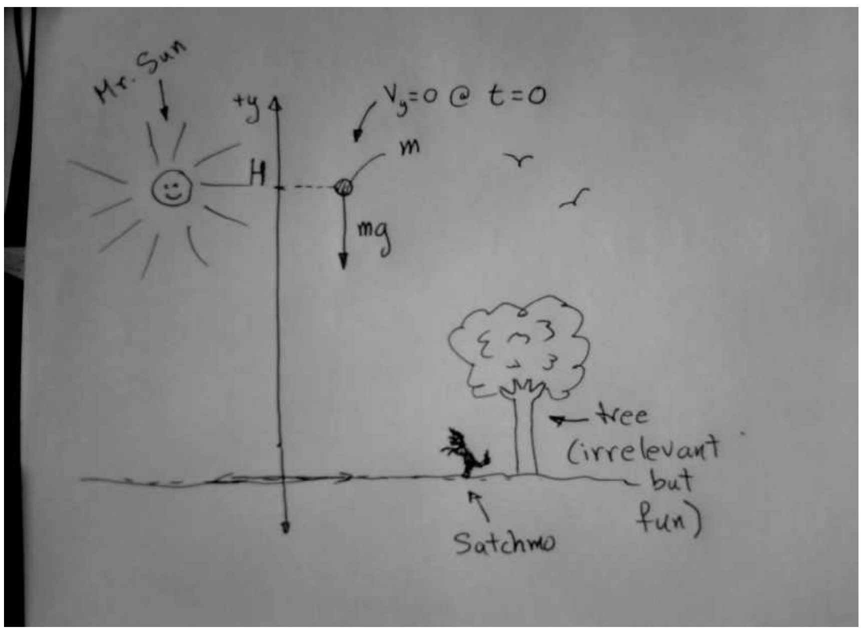


Figure 1.10: A picture of a ball being dropped from a height H, with a suitable one-dimensional coordinate system added. Note that the figure clearly indicates that it is the *force of gravity* that makes it fall. The pictures of Satchmo (my border collie) and the tree and sun and birds aren't strictly necessary and might even be distracting, but *my* right brain was bored when I drew this picture and they *do* orient the drawing and make it more fun!

Physics problems that you work and hand in that have no figure, no picture, not even additional hand-drawn decorations on a *provided* figure will rather soon ***lose points in the grading scheme!*** At first we (the course faculty) might just remind you and not take points off, but by your second assignment you'd better be adding *some* relevant artwork to every solution[57]. Figure 1.10 is what an actual figure you might draw to accompany a problem might

[56]At least for the next couple of weeks... but seriously, this rubric is useful all the way up to *graduate* physics.

[57]This has two benefits – one is that it actually is a critical step in solving the problem, the other is that drawing

look like.

Note well a couple of things about this figure. First of all, it is *large* – it took up 1/4+ of the unlined/white page I drew it on. This is actually good practice – ***do not draw postage-stamp sized figures!*** Draw them large enough that you can decorate them, not with Satchmo but with things like coordinates, forces, components of forces, initial data reminders. This is your brain we're talking about here, because ***the paper is functioning as an extension of your brain*** when you *use* it and the pictures you draw on it and the algebra you do on it to help solve the problem. Is your brain postage-stamp sized? Don't worry about wasting paper – ***paper is cheap, physics educations are expensive***. Use a whole page (or more) of ***plain white*** (e.g. copier) *paper* per problem solution at this point, ***not*** three problems per page with figures that require a magnifying glass to make out on lined paper ripped out of a spiral notebook!

When I (or your instructor) solve problems *with* you, this is the kind of thing you'll see us draw, over and over again, on the board, on paper at a table, wherever. In time, physicists become pretty good schematic artists and so should you. However, in a *textbook* we want things to be clearer and prettier, so I'll redraw this in figure 1.11, this time with a computer drawing tool (xfig) that I'll use for drawing *most* of the figures included in the textbook. Alas, it won't have Satchmo, but it does have *all of the important stuff* that should be on your hand-drawn figures.

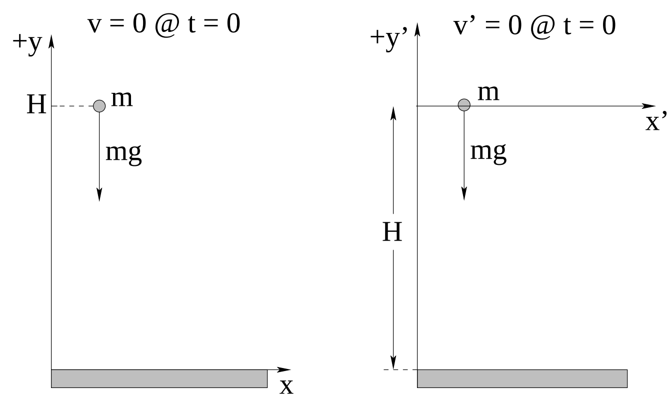


Figure 1.11: The same figure and coordinate system, drawn "perfectly" with xfig, plus a *second* (alternative) coordinatization.

Note that I drew *two* alternative ways of adding coordinates to the problem. The x-y coordinate system on the left is appropriate if you visualize the problem from the ground, looking

engages the right hemisphere of your brain (the left hemisphere is the one that does the algebra). The right hemisphere is the one that controls formation of long term memory, and it can literally get *bored, independently* of the left hemisphere and interrupt your ability to work. If you've ever worked for a very long time on writing something very dry (left hemisphere) or doing lots of algebraic problems (left hemisphere) and found your eyes being almost irresistably drawn up to look out the window at the grass and trees and ponies and bright sun, then know that it is your "right brain" that has taken over your body because it is dying in there, bored out of its (your!) gourd.

To keep the right brain happy while you do left brained stuff, give it something to do – listen to music, draw pictures or visualize a lot, take five minute right-brain-breaks and *deliberately* look at something visually pleasing. If your right brain is happy, you can work longer and better. If your right brain is *engaged in solving the problem* you will remember what you are working on much better, it will make more sense, and your attention won't wander as much. Physics is a *whole brain subject*, and the more pathways you use while working on it, the easier it is to understand and remember!

up like Satchmo, where the ground is at zero height. This might be e.g. dropping a ball off of the top of Duke Chapel, for example, with you on the ground watching it fall.

The x'-y' coordinate system on the right works if you visualize the problem as something like dropping the same ball into a well, where the ground is still at "zero height" but now it falls down to a *negative* height H from zero instead of starting at H and falling to height zero. Or, *you* dropping the ball from the top of the Duke Chapel and counting "$y' = 0$ as the height where *you* are up there (and the initial position in y' of the ball), with the ground at $y' = -H$ below the final position of the ball after it falls.

Now pay attention, because this is important: **Physics doesn't care which coordinate system you use!** Both of these coordinatizations of the problem are inertial reference frames. If you think about it, you will be able to see how to transform the answers obtained in one coordinate system into the corresponding answers in the other (basically subtract a constant H from the values of y in the left hand figure and you get y' in the right hand figure, right?). Newton's Laws will work perfectly in either inertial reference frame[58], and truthfully there are an *infinite* number of coordinate frames you could choose that would all describe the same problem in the end. You can therefore *choose the frame that makes the problem easiest to solve!*

Personally, from experience I prefer the left hand frame – it makes the algebra a tiny bit prettier – but the one on the right is really almost as good. I reject without thinking about it all of the frames where the mass m e.g. starts at the initial position $y_i = H/2$ and falls down to the final position $y_f = -H/2$. I *do* sometimes consider a frame like the one on the right with y positive pointing *down*, but it often bothers students to have "down" be positive (even though it is very natural to orient our coordinates so that $\vec{F}$ points in the positive direction of one of them) so we'll work into that gently. Finally, I did draw the x (horizontal) coordinate and ignored altogether for now the z coordinate that in principle is pointing out of the page in a right-handed coordinate frame. We don't really need either of these because no aspect of the motion will change x or z (there are *no forces acting* in those directions) so that the problem is effectively one-dimensional.

Next, we have to put in the *physics*, which at this point means: **Draw in all of the forces that act on the mass as proportionate vector arrows in the direction of the force.** The "proportionate" part will be difficult at first until you get a feeling for how large the forces are likely to be relative to one another but in this case there is only *one* force, gravity that acts, so we can write on our page (and on our diagram) the *vector* relation:

$$\vec{F} = -mg\hat{y} \tag{1.45}$$

or if you prefer, you can write the dimension-labelled scalar equation for the magnitude of the force in the y-direction:

$$F_y = -mg \tag{1.46}$$

Note well! Either of these is acceptable **vector notation** because the force is a **vector** (magnitude and direction). So is the decoration on the figure – an arrow for direction labelled mg.

What **is not quite right** (to the tune of minus a point or two at the discretion of the grader) is to just write $F = mg$ on your paper without indicating its direction *somehow*. Yes, this

[58]For the moment you can take my word for this, but we will *prove* it in the next week/chapter when we learn how to systematically change between coordinate frames!

is the magnitude of the force, but in what direction does it point in the particular coordinate system you drew into your figure? After all, you could have made $+x$ point down as easily as $-y$! Practice connecting your visualization of the problem in the coordinates you selected to a correct algebraic/symbolic description of the *vectors* involved.

In context, we don't really need to write $F_x = F_z = 0$ because they are so clearly irrelevant. However, in many other problems we will need to include either or both of these. You'll quickly get a feel for when you do or don't need to worry about them – a reasonable "rule" for this is represented in the figure above – the particle has no x velocity, there are no forces *at all* in the x-direction, and we could even make the initial x coordinate of the particle zero. Nothing happens that is at all interesting in the x direction, so we more or less ignore it.

Now comes the *key step* – setting up all of the algebra that leads to the solution. We write *Newton's Second Law for the mass* m, and *algebraically solve for the acceleration!* Since there is only one relevant component of the force in this one-dimensional problem, we only need to do this one time for the *scalar* equation for that component.:

$$\begin{aligned} F_y = -mg &= ma_y \\ ma_y &= -mg \\ a_y &= -g \\ \frac{d^2y}{dt^2} &= \frac{dv_y}{dt} = -g \end{aligned} \tag{1.47}$$

where $g = 10$ m/second^2 is the *constant* (within 2%, close to the Earth's surface, remember).

We are all but done at this point. The last line (the algebraic expression for the acceleration) is called the *equation of motion* for the system, and one of our chores will be to learn how to solve several common kinds of equation of motion. This one is a **constant acceleration problem**. Let's do it.

Here is the algebra involved. **Learn it.** Practice doing this until it is second nature when solving simple problems like this. I do *not recommend* memorizing the solution you obtain at the end, even though when you have solved the problem enough times you will probably remember it anyway for the rest of your share of eternity. Start with the equation of motion for a constant acceleration:

$$\begin{aligned} \frac{dv_y}{dt} &= -g \qquad \text{Next, multiply both sides by } dt \text{ to get:} \\ dv_y &= -g\ dt \qquad \text{Then integrate both sides:} \\ \int dv_y &= -\int g\ dt \qquad \text{doing the indefinite integrals to get:} \\ v_y(t) &= -gt + C \end{aligned} \tag{1.48}$$

The final C is the *constant of integration* of the indefinite integrals. We have to evaluate it using the *given* (usually initial) *conditions*. In this case we know that:

$$v_y(0) = -g \cdot 0 + C = C = 0 \tag{1.49}$$

(Recall that we even drew this into our figure to help remind us – it is the bit about being "dropped from rest" at time $t = 0$.) Thus:

$$v_y(t) = -gt \tag{1.50}$$

We now know the *velocity* of the dropped ball as a function of time! This is good, we are likely to need it. However, the *solution* to the dynamical problem is the trajectory function, $y(t)$. To find it, we repeat the same process, but now use the definition for v_y in terms of y:

$$\begin{aligned}\frac{dy}{dt} &= v_y(t) = -gt \qquad \text{Multiply both sides by } dt \text{ to get:}\\ dy &= -gt\,dt \qquad \text{Next, integrate both sides:}\\ \int dy &= -\int gt\,dt \qquad \text{to get:}\\ y(t) &= -\frac{1}{2}gt^2 + D \end{aligned} \tag{1.51}$$

The final D is *again* the constant of integration of the indefinite integrals. We *again* have to evaluate it using the given (initial) conditions in the problem. In this case we know that:

$$y(0) = -\frac{1}{2}g \cdot 0^2 + D = D = H \tag{1.52}$$

because we dropped it from an initial height $y(0) = H$. Thus:

$$y(t) = -\frac{1}{2}gt^2 + H \tag{1.53}$$

and we know *everything there is to know about the motion!* We know in particular exactly where it is at all times (until it hits the ground) as well as how fast it is going and in what direction. Sure, later we'll learn how to evaluate other quantities that *depend* on these two, but with the solutions in hand evaluating those quantities will be (we hope) trivial.

Finally, we have to **answer any questions that the problem might ask!** Note well that the problem *may not have told you* to evaluate $y(t)$ and $v_y(t)$, but in many cases you'll need them anyway to answer the questions they *do* ask. Here are a couple of common questions you can now answer using the solutions you just obtained:

a) How *long* will it take for the ball to reach the ground?

b) How *fast* is it going when it reaches the ground?

To answer the first one, we use a bit of algebra. "The ground" is (recall) $y = 0$ and it will reach there at some specific time (the time we want to solve for) t_g. We write the condition that it is at the ground at time t_g:

$$y(t_g) = -\frac{1}{2}gt_g^2 + H = 0 \tag{1.54}$$

If we rearrange this and solve for t_g we get:

$$t_g = \pm\sqrt{\frac{2H}{g}} \tag{1.55}$$

Hmmm, there seem to be *two* times at which $y(t_g)$ equals zero, one in the past and one in the future. The right answer, of course, must be the one in the future: $t_g = +\sqrt{2H/g}$, but you should think about what the one in the past *means*, and how the *algebraic* solution we've just developed is ignorant of things like your hand holding the ball before $t = 0$ and just what value of y corresponds to "the ground"...

That was pretty easy. To find the speed at which it hits the ground, one can just take our correct (future) time and plug it into v_y! That is:

$$v_g = v_y(t_g) = -gt_g = -g\sqrt{\frac{2H}{g}} = -\sqrt{2gH} \tag{1.56}$$

Note well that it is going *down* (in the negative y direction) when it hits the ground. This is a good hint for the previous puzzle. What direction would it have been going at the negative time? What kind of motion does the overall solution describe, on the interval from $t = (-\infty, \infty)$? Do we need to use a certain amount of *common sense* to avoid using the algebraic solution for times or values of y for which they *make* no sense, such as $y < 0$ or $t < 0$ (in the ground or before we let go of the ball, respectively)?

The last thing we might look at I'm going to let you do on your own (don't worry, it's easy enough to do in your head). *Assuming* that this algebraic solution is valid for any reasonable H, how fast does the ball hit the ground after falling (say) 5 meters? How about $20 = 4 * 5$ meters? How about $80 = 16 * 5$ meters? How long does it take for the ball to fall 5 meters, 20 meters, 80 meters, etc? In this course we won't do a *lot* of arithmetic, but whenever we learn a new idea with parameters like g in it, it is useful to do a *little* arithmetical exploration to see what a "reasonable" answer looks like. Especially note how the answers *scale* with the height – if one drops it from 4x the height, how much does that increase the time it falls and speed with which it hits?

One of these heights causes it to hit the ground in one second, and all of the other answers scale with it like the square root. If you happen to remember this height, you can actually estimate how long it takes for a ball to fall almost *any* height in your head with a division and a square root, and if you multiply the time answer by ten, well, there is the speed with which it hits! We'll do some conceptual problems that help you understand this *scaling* idea for homework.

This (a falling object) is nearly a perfect problem archetype or example for one dimensional motion. Sure, we can make it more complicated, but usually we'll do that by having *more than one thing* move in one dimension and then have to figure out how to solve the two problems *simultaneously* and answer questions given the results.

Let's take a short break to formally solve the equation of motion we get for a constant force in one dimension, as the general solution exhibits two constants of integration that we need to be able to identify and evaluate from initial conditions. Note well that the next problem is almost identical to the former one. It just differs in that you are given the force $\vec{F}$ itself, not a knowledge that the force is e.g. "gravity".

Example 1.7.2: A Constant Force in One Dimension

This time we'll imagine a different problem. A car of mass m is travelling at a constant speed v_0 as it enters a long, nearly straight merge lane. A distance d from the entrance, the driver presses the accelerator and the engine exerts a constant force of magnitude F on the car.

a) How long does it take the car to reach a final velocity $v_f > v_0$?

b) How far (from the entrance) does it travel in that time?

As before, we need to start with a **good picture** of what is going on. Hence a car:

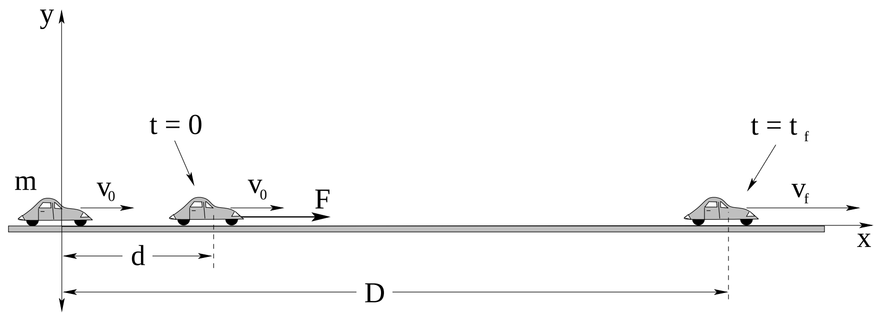


Figure 1.12: *One* possible way to portray the motion of the car and coordinatize it.

In figure 1.12 we see what we can imagine are three "slices" of the car's position as a function of time at the moments described in the problem. On the far left we see it "entering a long, nearly straight merge lane". The second position corresponds to the time the car is a distance d from the entrance, which is also the time the car starts to accelerate because of the force F. I chose to start the clock then, so that I can integrate to find the position as a function of time while the force is being applied. The final position corresponds to when the car has had the force applied for a time t_f and has acquired a velocity v_f. I labelled the distance of the car from the entrance D at that time. The mass of the car is indicated as well.

This figure completely captures the important features of the problem! Well, almost. There are two forces I ignored altogether. One of them is *gravity*, which is pulling the car *down*. The other is the so-called *normal force* exerted by the road on the car – this force pushes the car *up*. I ignored them because my experience and common sense tell me that under ordinary circumstances, on level ground, the road doesn't push on the car so that it jumps into the air, nor does gravity pull the car down into the road – the two forces will *balance* and the car will not move or accelerate in the vertical direction. In a bit, we'll take these forces into explicit account too as sometimes they do *not* cancel, but here I'm just going to use my intuition that they will cancel and hence that the y-direction can be ignored. In this case, all of the motion is going to be nicely one dimensional in the x-direction as I've defined it with my coordinate axes.

It's time to follow our ritual. We will write Newton's Second Law and solve for the acceleration (obtaining an equation of motion). Then we will integrate twice to find first $v_x(t)$ and then $x(t)$. We will have to be extra careful with the constants of integration this time, and in fact will get a very general solution, one that can be applied to all constant acceleration problems, although I *do not recommend* that you *memorize* this solution and try to use it every time you see Newton's Second Law! For one thing, we'll have quite a few problems this year where the force, and acceleration, are *not constant* and in those problems the solution we will derive is *wrong*. Alas, to my own extensive and direct experience, students that memorize kinematic solutions to the constant acceleration problem instead of learning to solve it with actual integration done *every time* almost invariably try applying the solution to e.g. the harmonic oscillator problem later, and I hate that. So don't memorize the answer; learn how to derive it and practice the derivation until (sure) you *know* the result, and also *know* when you can use it.

Thus:

$$\begin{aligned} F &= ma_x \\ a_x &= \frac{F}{m} = a_0 \quad \text{(a constant)} \\ \frac{dv_x}{dt} &= a_0 \end{aligned} \tag{1.57}$$

Next, multiply through by dt and integrate both sides:

$$v_x(t) = \int dv_x = \int a_0\ dt = a_0 t + V = \frac{F}{m}t + V \tag{1.58}$$

Either of the last two are valid answers, provided that we define $a_0 = F/m$ somewhere in the solution and also provided that the problem doesn't explicitly ask for an answer to be given in terms of F and m. V is a constant of integration that we will evaluate below.

Note that if $a_0 = F/m$ was *not* a constant (say that F(t) is a function of time) then we would have to *do the integral*:

$$v_x(t) = \int \frac{F(t)}{m}\ dt = \frac{1}{m}\int F(t)\ dt = ??? \tag{1.59}$$

At the very least, we would have to know the explicit functional form of $F(t)$ to proceed, and the answer would *not* be linear in time.

At time $t = 0$, the velocity of the car in the x-direction is v_0, so (check for yourself) $V = v_0$ and:

$$v_x(t) = a_0 t + v_0 = \frac{dx}{dt} \tag{1.60}$$

We multiply *this* equation by dt on both sides, integrate, and get:

$$x(t) = \int dx = \int (a_0 t + v_0)\ dt = \frac{1}{2}a_0 t^2 + v_0 t + x_0 \tag{1.61}$$

where x_0 is the constant of integration. We note that at time $t = 0$, $x(0) = d$, so $x_0 = d$. Thus:

$$x(t) = \frac{1}{2}a_0 t^2 + v_0 t + d \tag{1.62}$$

It is worth collecting the two basic solutions in one place. It should be obvious that for *any* one-dimensional (say, in the x-direction) constant acceleration $a_x = a_0$ problem we will *always* find that:

$$v_x(t) = a_0 t + v_0 \tag{1.63}$$

$$x(t) = \frac{1}{2}a_0 t^2 + v_0 t + x_0 \tag{1.64}$$

where x_0 is the x-position at time $t = 0$ and v_0 is the x-velocity at time $t = 0$. You can see why it is so very tempting to just memorize this result and pretend that you know a piece of physics, but don't!

The algebra that led to this answer is basically ordinary math *with units*. As we've seen, "math with units" has a special name all its own – *kinematics* – and the pair of equations 1.63

and 1.64 are called the *kinematic solutions to the constant acceleration problem*. Kinematics should be contrasted with *dynamics*, the physics of forces and laws of nature that lead us to equations of motion. One way of viewing our solution strategy is that – after drawing and decorating our figure, of course – we solve first the *dynamics problem* of writing our dynamical principle (Newton's Second Law with the appropriate vector total force), turning it into a differential equation of motion, then solving the resulting *kinematics problem* represented by the equation of motion with *calculus*. Don't be tempted to *skip* the calculus and try to memorize the kinematic solutions – it is just as important to understand and be able to do the kinematic calculus quickly and painlessly as it is to be able to set up the dynamical part of the solution.

Now, of course, we have to actually answer the *questions* given above. To do this requires as before logic, common sense, intuition, experience, and math. First, at what *time* t_f does the car have speed v_f? When:

$$v_x(t_f) = v_f = a_0 t_f + v_0 \tag{1.65}$$

of course. You can easily *solve* this for t_f. Note that I just transformed the English statement "At t_f, the car must have speed v_f" into an algebraic equation that means the *exact same thing!*

Second, what is D? Well in *English*, the distance D from the entrance is where the car is at time t_f, when it is also travelling at speed v_f. If we turn this sentence into an equation we get:

$$x(t_f) = D = \frac{1}{2} a_0 t_f^2 + v_0 t_f + d \tag{1.66}$$

Again, having solved the previous equation algebraically, you can substitute the result for t_f into this equation and get D in terms of the originally given quantities! The problem is solved, the questions are answered, we're finished.

Or rather, *you* will be finished, after you fill in these last couple of steps on your own!

1.7.1: Solving Problems with More Than One Object

One of the keys to answering the questions in both of these examples has been turning easy-enough statements in English into equations, and then solving the equations to obtain an answer to a question also framed in English. So far, we have solved only single equations, but we will *often* be working with more than one thing at a time, or combining two or more principles, so that we have to solve several *simultaneous* equations.

The only change we might make to our existing solution strategy is to construct and solve the equations of motion for *each* object or independent aspect (such as dimension) of the problem. In a moment, we'll consider problems of the latter sort, where this strategy will work when the *force in one coordinate direction is independent of the force in another coordinate direction!* . First, though, let's do a couple of very simple one-dimensional problems with *two* objects with some sort of constraint connecting the motion of one to the motion of the other.

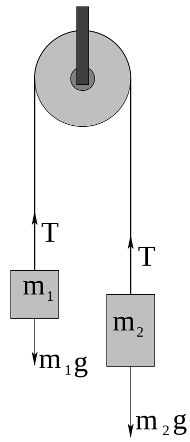


Figure 1.13: Atwood's Machines consists of a pair of masses (usually of different mass) connected by a string that runs over a pulley as shown. Initially we idealize by considering the pulley to be massless and frictionless, the string to be massless and unstretchable, and ignore drag forces.

Example 1.7.3: Atwood's Machine

A mass m_1 and a second mass m_2 are hung at both ends of a massless, unstretchable string that runs over a frictionless, massless pulley as shown in figure 1.13. Gravity near the Earth's surface pulls both down. Assuming that the masses are released from rest at time $t = 0$, find:

a) The acceleration of both masses;

b) The tension T in the string;

c) The speed of the masses after they have moved through a distance H in the direction of the more massive one.

The trick of this problem is to note that if mass m_2 goes down by a distance (say) x, mass m_1 goes up by the *same* distance x and vice versa. The magnitude of the displacement of one is the same as that of the other, as they are connected by a taut unstretchable string. This also means that the speed of one rising equals the speed of the other falling, the magnitude of the acceleration of one up equals the magnitude of the acceleration of the other down. So even though it at first *looks* like you need two coordinate systems for this problem, x_1 (measured from m_1's initial position, up or down) will equal x_2 (measured from m_2's initial position, down or up) be the same. We therefore can just use x to describe this displacement (the displacement of m_1 up and m_2 down from its starting position), with v_x and a_x being the same for both masses with the same convention.

This, then, is a *wraparound* one-dimensional coordinate system, one that "curves around the pulley". In these coordinates, Newton's Second Law for the two masses becomes the two equations:

$$F_1 = T - m_1 g = m_1 a_x \tag{1.67}$$

$$F_2 = m_2 g - T = m_2 a_x \tag{1.68}$$

This is a set of two equations and two unknowns (T and a_x). It is easiest to solve by elimination. If we add the two equations we eliminate T and get:

$$m_2 g - m_1 g = (m_2 - m_1)g = m_1 a_x + m_2 a_x = (m_1 + m_2)a_x \tag{1.69}$$

or

$$a_x = \frac{m_2 - m_1}{m_1 + m_2} g \tag{1.70}$$

In the figure above, if $m_2 > m_1$ (as the figure suggests) then *both* mass m_2 will accelerate down *and* m_1 will accelerate up with this constant acceleration.

We can find T by substituting this value for a_x into *either* force equation:

$$\begin{aligned}
T - m_1 g &= m_1 a_x \\
T - m_1 g &= \frac{m_2 - m_1}{m_1 + m_2} m_1 g \\
T &= \frac{m_2 - m_1}{m_1 + m_2} m_1 g + m_1 g \\
T &= \frac{m_2 - m_1}{m_1 + m_2} m_1 g + \frac{m_2 + m_1}{m_1 + m_2} m_1 g \\
T &= \frac{2 m_2 m_1}{m_1 + m_2} g
\end{aligned} \tag{1.71}$$

a_x is constant, so we can evaluate $v_x(t)$ and $x(t)$ exactly as we did for a falling ball:

$$\begin{aligned}
a_x = \frac{dv_x}{dt} &= \frac{m_2 - m_1}{m_1 + m_2} g \\
dv_x &= \frac{m_2 - m_1}{m_1 + m_2} g\ dt \\
\int dv_x &= \int \frac{m_2 - m_1}{m_1 + m_2} g\ dt \\
v_x &= \frac{m_2 - m_1}{m_1 + m_2} gt + C \\
v_x(t) &= \frac{m_2 - m_1}{m_1 + m_2} gt
\end{aligned} \tag{1.72}$$

and then:

$$\begin{aligned}
v_x(t) = \frac{dx}{dt} &= \frac{m_2 - m_1}{m_1 + m_2} gt \\
dx &= \frac{m_2 - m_1}{m_1 + m_2} gt\ dt \\
\int dx &= \int \frac{m_2 - m_1}{m_1 + m_2} gt\ dt \\
x &= \frac{1}{2} \frac{m_2 - m_1}{m_1 + m_2} gt^2 + C' \\
x(t) &= \frac{1}{2} \frac{m_2 - m_1}{m_1 + m_2} gt^2
\end{aligned} \tag{1.73}$$

(where C and C' are set from our knowledge of the initial conditions, $x(0) = 0$ and $v(0) = 0$ in the coordinates we chose).

Now suppose that the blocks "fall" a height H (only m_2 actually falls, m_1 goes up). Then we can, as before, find out how long it takes for $x(t_h) = H$, then substitute this into $v_x(t_h)$ to find the speed. I leave it as an exercise to show that this answer is:

$$v_x(t_h) = \sqrt{\left(\frac{m_2 - m_1}{m_1 + m_2}\right) 2gH} \tag{1.74}$$

Example 1.7.4: Braking for Bikes, or Just Breaking Bikes?

A car of mass M is overtaking a bicyclist. Initially, the car is travelling at speed v_{0c} and the bicyclist is travelling at $v_{0b} < v_{0c}$ in the same direction. At a time that the bicyclist is D meters away, the driver of the car suddenly sees that he is on a collision course and applies the brakes, exerting a force $-F$ on his car (where the minus sign just means that he is slowing down, diminishing his velocity.

Assuming that the bicyclist doesn't speed up or slow down, does he hit the bike?

At this point you should have a pretty good idea how to proceed for *each* object. First, we'll draw a figure with both objects and formulate the equations of motion for each object separately. Second, we'll solve the equations of motion for reach object. Third, we'll write an equation that captures the *condition* that the car hits the bike, and see if that equation has any solutions. If so, then it is likely that the car will be breaking, not braking (in time)!

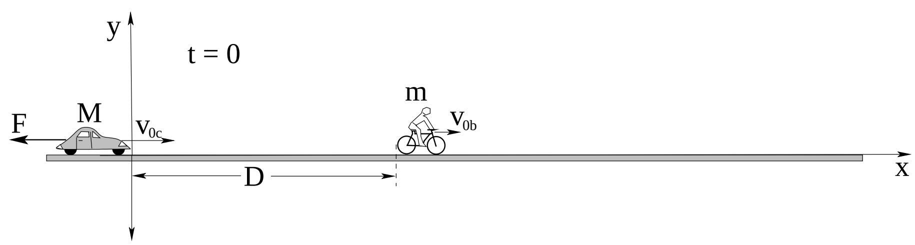


Figure 1.14: The initial picture of the car overtaking the bike at the instant it starts to brake. Again we will ignore the forces in the y-direction as we know that the car doesn't *jump over* the bike and we'll pretend that the biker can't just turn and get out of the way as well.

Here's the solution, *without* most of the details. You should work through this example, filling in the missing details and making the solution all pretty. The magnitude of the acceleration of the car is $a_c = F/M$, and we'll go ahead and use this constant acceleration a_c to formulate the answer – we can always do the arithmetic and substitute at the end, given some particular values for F and M.

Integrating this (and using $x_c(0) = 0$, $v_c(0) = v_{0c}$) you will get:

$$v_c(t) = -a_c t + v_{0c} \tag{1.75}$$

$$x_c(t) = -\frac{1}{2} a_c t^2 + v_{0c} t \tag{1.76}$$

The acceleration of the bike is $a_b = 0$. This means that:

$$v_b(t) = a_b t + v_{0b} = v_{0b} \tag{1.77}$$

The velocity of the bike is *constant* because there is no (net) force acting on it and hence it has no acceleration. Integrating this one gets (using $x_b(0) = D$):

$$x_b(t) = v_{0b}t + D \tag{1.78}$$

Now the big question: Does the car hit the bike? If it *does*, it does so at some *real time*, call it t_h. "Hitting" means that there is no distance between them – they are at the same place at the same time, in particular at *this* time t_h. Turning this sentence into an equation, the condition for a collision is algebraically:

$$x_b(t_h) = v_{0b}t_h + D = -\frac{1}{2}a_c t_h^2 + v_{0c}t_h = x_c(t_h) \tag{1.79}$$

Rearranged, this is a *quadratic equation*:

$$\frac{1}{2}a_c t_h^2 - (v_{0c} - v_{0b})\, t_h + D = 0 \tag{1.80}$$

and therefore has two roots. If we write down the quadratic formula:

$$t_h = \frac{(v_{0c} - v_{0b}) \pm \sqrt{(v_{0c} - v_{0b})^2 - 2a_c D}}{a_c} \tag{1.81}$$

we can see that there will only be a *real* (as opposed to *imaginary*) time t_h that solves the collision condition if the argument of the square root is non-negative. That is:

$$(v_{0c} - v_{0b})^2 \geq 2a_c D \tag{1.82}$$

If this is true, there will be a "collision" at times corresponding to the two real roots to the quadratic formula[59]. If it is false, the car will never reach the bike.

There is actually a second way to arrive at this result. One can find the time t_s that the car is travelling at the same *speed* as the bike. That's really pretty easy:

$$v_{0b} = v_c(t_s) = -a_c t_s + v_{0c} \tag{1.83}$$

or

$$t_s = \frac{(v_{0c} - v_{0b})}{a_c} \tag{1.84}$$

Now we locate the car relative to the bike. If the collision *hasn't* happened by t_s it never will, as afterwards the car will be slower than the bike and the bike will pull away. If the position of the car is behind (or barely equal to) the position of the bike at t_s, all is well, no collision occurs. That is:

$$x_c(t_s) = -\frac{1}{2}a_c t_s^2 + v_{0c}t_s \leq v_{0b}t_s + D \tag{1.85}$$

if no collision occurs. It's left as an exercise to show that this leads to the same condition that the quadratic gives you.

Next, let's see what happens when we have only one object but motion in two dimensions.

---

[59]Two roots? Yes, amusingly enough our mathematical solution assumes that the instant the car comes to rest, it goes into reverse and *continues to accelerate in the negative direction*. It therefore goes *back and runs over the hapless biker a second time*. Assuming also, of course, that the biker somehow continues going forward at the same speed after being hit.

The moral of the story is: *Remember that the mathematical solutions can easily have non-physical, irrelevant, or just plain silly parts!* And yeah, we might well take off a point or two if you put down one of these silly solutions in place of or addition to the correct one, without any sort of comment...

1.8: Motion in Two Dimensions

The idea of motion in two or more dimensions is very simple. Force is a *vector*, and so is acceleration. Newton's Second Law is a recipe for taking the total force and converting it into a differential equation of motion:

$$\vec{a} = \frac{d^2\vec{r}}{dt^2} = \frac{\vec{F}_{\rm tot}}{m} \tag{1.86}$$

In the most general case, this can be quite difficult to solve. For example, consider the forces that act upon *you* throughout the day – every step you take, riding in a car, gravity, friction, even the wind exert forces subtle or profound on your mass and accelerate you first this way, then that as you move around. The total force acting on you varies wildly with time and place, so even though your trajectory *is* a solution to just such an equation of motion, computing it algebraically is out of the question. Computing it with a computer would be straightforward *if* the forces were all known, but of course they vary according to your volition and the circumstances of the moment and are hardly knowable ahead of time.

However, *much* of what happens in the world around you can actually be at least approximated by relatively simple (if somewhat idealized) models and explicitly solved. These simple models generally arise when the forces acting are due to the "well-known" forces of nature or force rules listed above and hence point in specific directions (so that their vector description can be analyzed) and are either constant in time or vary in some known way so that the calculus of the solution is tractable[60].

We will now consider only these latter sorts of forces: forces that act in a well-defined direction with a computable value (initially, with a computable *constant* value, or a value that varies in some simple way with position or time). If we write the equation of motion out in components:

$$a_x = \frac{d^2x}{dt^2} = \frac{F_{{\rm tot},x}}{m} \tag{1.87}$$

$$a_y = \frac{d^2y}{dt^2} = \frac{F_{{\rm tot},y}}{m} \tag{1.88}$$

$$a_z = \frac{d^2z}{dt^2} = \frac{F_{{\rm tot},z}}{m} \tag{1.89}$$

we will often reduce the complexity of the problem from a "three dimensional problem" to three "one dimensional problems" of the sort we just learned to solve in the section above.

Of course, when this is possible, there's a trick to it. The trick is this:

> **Select a coordinate system in which one or more of the coordinate axes are in a direction where we *know the acceleration*, for example a direction where it is *zero* (straight line motion) or v^2/r (circular motion) and hence where we instantly know the solution.**

In many cases this means we should:

[60]"Tractable" here means that it can either be solved algebraically, true for many of the force laws or rules, or at least solved numerically. In this course you may or may not be required or expected to explore numerical solutions to the differential equations with e.g. matlab, octave, or mathematica.

> **Select a coordinate system in which one of the coordinate axes is aligned with the total force.**

as this means that the force in the other two directions *is* zero, hence the acceleration is zero, hence the motion in those directions is (hopefully) "simple", as in constant straight line motion or no motion at all.

We won't *always* be able to do this, but when it can it will get us off to a very good start, and *trying* it will help us understand what to do when we hit problems where this alone won't quite work or help us solve the problem.

Again, the reason this step (when possible) simplifies the problem is simple enough to understand: In this particular coordinate frame (with the total force pointing in a single direction along one of the coordinate axes), the total force in the *other* directions adds up to zero! That means that all *acceleration* occurs only along the selected coordinate direction. Solving the equations of motion in the other directions is then trivial – it is motion with a constant velocity (which may be zero, as in the case of dropping a ball vertically down from the top of a tower in the problems above, or not in the case of ballistic trajectories examined below). Solving the equation of motion in the direction of the total force itself is then "the problem", and you will need lots of practice and a few good examples to show you how to go about it.

To make life even simpler while we are learning, we will now further restrict ourselves to the class of problems where the acceleration *and* velocity in one of the three dimensions is zero. In that case the value of that coordinate is constant, and may as well be taken to be zero. The motion (if any) then occurs in the remaining two dimensional *plane* that contains the origin. In the problems below, we will find it useful to use one of *two possible* two-dimensional coordinate systems to solve for the motion: Cartesian coordinates (which we've already begun to use, at least in a trivial way) and Plane Polar coordinates, which we will review in context below.

As you will see, solving problems in two or three dimensions with a constant force direction simply introduces a few extra steps into the solution process:

- Decomposing the known forces into a coordinate system where one of the coordinate axes lines up with the (expected) total force...
- Solving the individual one-dimensional motion problems (where one or two of the resulting solutions will usually be "trivial", e.g. constant motion)...
- Finally, reconstructing the overall (vector) solution from the individual solutions for the independent vector coordinate directions...

and answering any questions as usual.

1.8.1: Free Flight Trajectories – Projectile Motion

Perhaps the simplest example of this process adds just one small change to our first example. Instead of dropping a particle *straight down* let us imagine *throwing* the ball off of a tower, or *firing a cannon*, or *driving a golf ball* off of a tee or *shooting a basketball*. All of these are examples of *projectile motion* – motion under the primary action of vertical gravity where the initial velocity in some horizontal direction is *not* zero.

Note well that we will necessarily idealize our treatment by (initially) neglecting some of the many things that might affect the trajectory of all of these objects in the real world – drag forces which both slow down e.g. a golf ball and exert "lift" on it that can cause it to hook or slice, the fact that the earth is not really an inertial reference frame and is rotating out underneath the free flight trajectory of a cannonball, creating an apparent deflection of actual projectiles fired by e.g. naval cannons. That is, *only* gravity near the earth's surface will act on our ideal particles for now.

The easiest way to teach you how to handle problems of this sort is just to do a few examples – there are really only three distinct cases one can treat – two rather special ones and the general solution. Let's start with the simplest of the special ones.

Example 1.8.1: Trajectory of a Cannonball

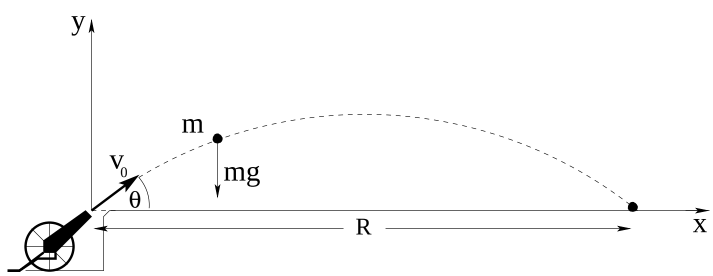


Figure 1.15: An idealized cannon, neglecting the drag force of the air. Let x be the horizontal direction and y be the vertical direction, as shown. Note well that $\vec{F}_g = -mg\hat{y}$ points along one of the coordinate directions while $F_x = (F_z =)0$ in this coordinate frame.

A cannon fires a cannonball of mass m at an initial speed v_0 at an angle θ with respect to the ground as shown in figure 1.15. Find:

a) The time the cannonball is in the air.

b) The range of the cannonball.

We've already done the first step of a good solution – drawing a good figure, selecting and sketching in a coordinate system with one axis aligned with the total force, and drawing and labelling all of the forces (in this case, only one). We therefore proceed to write Newton's Second Law for *both* coordinate directions.

$$F_x = ma_x = 0 \tag{1.90}$$

$$F_y = ma_y = m\frac{d^2y}{dt^2} = -mg \tag{1.91}$$

We divide each of these equations by m to obtain two equations of motion, one for x and the other for y:

$$a_x = 0 \tag{1.92}$$
$$a_y = -g \tag{1.93}$$

We solve them independently. In x:

$$a_x = \frac{dv_x}{dt} = 0 \tag{1.94}$$

The derivative of any constant is zero, so the x-component of the velocity does not change in time. We find the initial (and hence constant) component using trigonometry:

$$v_x(t) = v_{0x} = v_0 \cos(\theta) \tag{1.95}$$

We then write *this* in terms of derivatives and solve it:

$$\begin{aligned}
v_x = \frac{dx}{dt} &= v_0 \cos(\theta) \\
dx &= v_0 \cos(\theta)\ dt \\
\int dx &= v_0 \cos(\theta) \int dt \\
x(t) &= v_0 \cos(\theta) t + C
\end{aligned}$$

We evaluate C (the constant of integration) from our knowledge that in the coordinate system we selected, $x(0) = 0$ so that $C = 0$. Thus:

$$x(t) = v_0 \cos(\theta) t \tag{1.96}$$

The solution in y is more or less identical to the solution that we obtained above dropping a ball, except the constants of integration are different:

$$\begin{aligned}
a_y = \frac{dv_y}{dt} &= -g \\
dv_y &= -g\ dt \\
\int dv_y &= -\int g\ dt \\
v_y(t) &= -gt + C'
\end{aligned} \tag{1.97}$$

For this problem, we know from trigonometry that:

$$v_y(0) = v_0 \sin(\theta) \tag{1.98}$$

so that $C' = v_0 \sin(\theta)$ and:

$$v_y(t) = -gt + v_0 \sin(\theta) \tag{1.99}$$

We write v_y in terms of the time derivative of y and integrate:

$$\begin{aligned}
\frac{dy}{dt} &= v_y(t) = -gt + v_0 \sin(\theta) \\
dy &= (-gt + v_0 \sin(\theta))\ dt \\
\int dy &= \int (-gt + v_0 \sin(\theta))\ dt \\
y(t) &= -\frac{1}{2} g t^2 + v_0 \sin(\theta) t + D
\end{aligned} \tag{1.100}$$

Again we use $y(0) = 0$ in the coordinate system we selected to set $D = 0$ and get:

$$y(t) = -\frac{1}{2}gt^2 + v_0 \sin(\theta)t \tag{1.101}$$

Collecting the results from above, our overall solution is thus:

$$x(t) = v_0 \cos(\theta)t \tag{1.102}$$
$$y(t) = -\frac{1}{2}gt^2 + v_0 \sin(\theta)t \tag{1.103}$$
$$v_x(t) = v_{0x} = v_0 \cos(\theta) \tag{1.104}$$
$$v_y(t) = -gt + v_0 \sin(\theta) \tag{1.105}$$

We know exactly where the cannonball is at all times, *and* we know exactly what its velocity is as well. Now let's see how we can answer the equations.

To find out how long the cannonball is in the air, we need to write an algebraic expression that we can use to identify when it hits the ground. As before (dropping a ball) "hitting the ground" in algebra-speak is $y(t_g) = 0$, so finding t_g such that this is true should do the trick:

$$\begin{aligned} y(t_g) = -\frac{1}{2}gt_g^2 + v_0 \sin(\theta)t_g &= 0 \\ \left(-\frac{1}{2}gt_g + v_0 \sin(\theta)\right) t_g &= 0 \end{aligned}$$

or

$$t_{g,1} = 0 \tag{1.106}$$
$$t_{g,2} = \frac{2v_0 \sin(\theta)}{g} \tag{1.107}$$

are the two roots of this (factorizable) quadratic. The first root obviously describes when the ball was fired, so it is the second one we want. The ball hits the ground after being in the air for a time

$$t_{g,2} = \frac{2v_0 \sin(\theta)}{g} \tag{1.108}$$

Now it is *easy* to find the range of the cannonball, R. R is just the value of $x(t)$ at the time that the cannonball hits!

$$R = x(t_{g,2}) = \frac{2v_0^2 \sin(\theta)\cos(\theta)}{g} \tag{1.109}$$

Using a trig identity one can also write this as:

$$R = \frac{v_0^2 \sin(2\theta)}{g} \tag{1.110}$$

The only reason to do this is so that one can see that the range of this projectile is *symmetric*: It is the same for $\theta = \pi/4 \pm \phi$ for any $\phi \in [0, \pi/4]$.

For your homework you will do a more general case of this, one where the cannonball (or golf ball, or arrow, or whatever) is fired off of the top of a cliff of height H. The solution will proceed *identically* except that the initial and final conditions may be different. In general, to find the time and range in this case one will have to solve a *quadratic equation* using the *quadratic formula* (instead of simple factorization) so if you haven't reviewed or remembered the quadratic formula before now in the course, please do so right away.

1.8.2: The Inclined Plane

The inclined plane is another archetypical problem for motion in two dimensions. It has many variants. We'll start with the *simplest* one, one that illustrates a new force, the *normal* force. Recall from above that the normal force is *whatever magnitude it needs to be* to prevent an object from moving *in* to a solid surface, and is always perpendicular (normal) to that surface in direction.

In addition, this problem beautifully illustrates the reason one selects coordinates aligned with the total force when that direction is consistent throughout a problem, if at all possible.

Example 1.8.2: The Inclined Plane

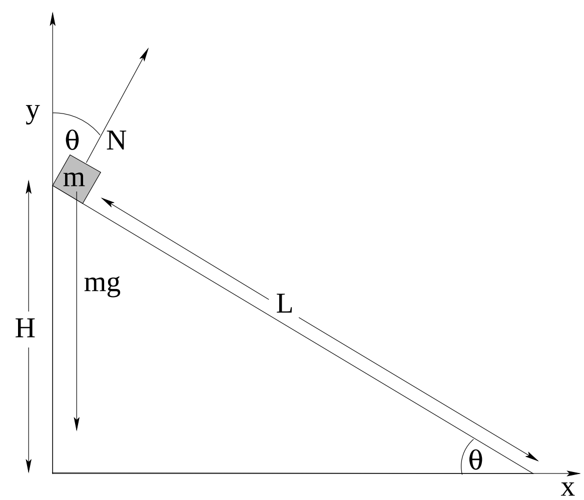


Figure 1.16: This is the naive/wrong coordinate system to use for the inclined plane problem. The problem *can* be solved in this coordinate frame, but the solution (as you can see) would be quite difficult.

A block m rests on a plane inclined at an angle of θ with respect to the horizontal. There is no friction (yet), but the plane exerts a normal force on the block that keeps it from falling straight down. At time $t = 0$ it is released (at a height $H = L\sin(\theta)$ above the ground), and we might then be asked any of the "usual" questions – how long does it take to reach the ground, how fast is it going when it gets there and so on.

The motion we expect is for the block to ***slide down the incline***, and for us to be able to solve the problem easily we have to use our *intuition* and ability to *visualize* this motion to select the best coordinate frame.

Let's start by doing the problem *foolishly*. Note well that in principle we actually *can* solve the problem set up this way, so it isn't really *wrong*, but in practice while *I* can solve it in this frame (having taught this course for 40 years and being pretty good at things like trig and calculus) it is somewhat less likely that *you* will have much luck if you haven't even used trig or taken a derivative for three or four years. Kids, Don't Try This at Home[61]...

[61]Or rather, by all means give it a try, especially after reviewing my solution.

In figure 1.16, I've drawn a coordinate frame that is lined up with gravity. However, gravity is not the *only* force acting any more. We expect the block to slide down the incline, not move straight down. We expect that the normal force will exert any force needed such that this is so. Let's see what happens when we try to decompose these forces in terms of our coordinate system.

We start by finding the components of $\vec{N}$, the vector normal force, in our coordinate frame:

$$N_x = N\sin(\theta) \tag{1.111}$$
$$N_y = N\cos(\theta) \tag{1.112}$$

where $N = |\vec{N}|$ is the (unknown) magnitude of the normal force.

We then add up the total forces in each direction and write Newton's Second Law for each direction's total force :

$$F_x = N\sin(\theta) = ma_x \tag{1.113}$$
$$F_y = N\cos(\theta) - mg = ma_y \tag{1.114}$$

Finally, we write our equations of motion for each direction:

$$a_x = \frac{N\sin(\theta)}{m} \tag{1.115}$$
$$a_y = \frac{N\cos(\theta) - mg}{m} \tag{1.116}$$

Unfortunately, we *cannot solve these two equations as written* yet. That is because we *do not know* the value of N; it is in fact something we need to *solve for!* To solve them we need to add a *condition* on the solution, expressed as an equation. The condition we need to add is that the motion is *down the incline*, that is, at all times:

$$\frac{y(t)}{L\cos(\theta) - x(t)} = \tan(\theta) \tag{1.117}$$

must be true as a constraint[62]. That means that:

$$\begin{aligned}
y(t) &= (L\cos(\theta) - x(t))\tan(\theta) \\
\frac{dy(t)}{dt} &= -\frac{dx(t)}{dt}\tan(\theta) \\
\frac{d^2y(t)}{dt^2} &= -\frac{d^2x(t)}{dt^2}\tan(\theta) \\
a_y &= -a_x\tan(\theta) \qquad (1.118)
\end{aligned}$$

where we used the fact that the time derivative of $L\cos(\theta)$ is zero! We can use this relation to eliminate (say) a_y from the equations above, solve for a_x, then backsubstitute to find a_y. Both are constant acceleration problems and hence we can easily enough solve them. But *yuk!* The solutions we get will be so very complicated (at least compared to choosing a better frame), with both x and y varying nontrivially with time.

Now let's see what happens when we choose the *right* (or at least a "good") coordinate frame according to the prescription given. Such a frame is drawn in 1.17:

[62]Note that the tangent involves the horizontal distance of the block *from the lower apex* of the inclined plane, $x' = L\cos(\theta) - x$ where x is measured, of course, from the origin.

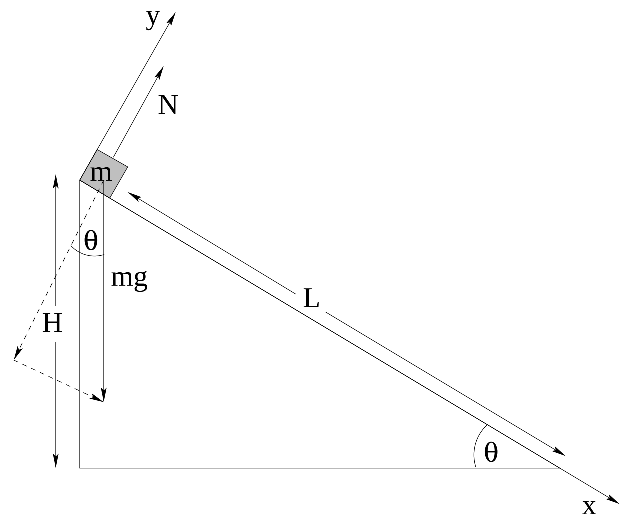


Figure 1.17: A good choice of coordinate frame has (say) the x-coordinate lined up with the total force and hence direction of motion.

As before, we can decompose the forces in this coordinate system, but now we need to find the components of the *gravitational* force as $\vec{N} = N\hat{y}$ is easy! Furthermore, ***we know that*** $a_y = 0$ ***and hence*** $F_y = 0$.

$$F_x = mg\sin(\theta) = ma_x \tag{1.119}$$

$$F_y = N - mg\cos(\theta) = ma_y = 0 \tag{1.120}$$

We can *immediately* solve the y equation for:

$$N = mg\cos(\theta) \tag{1.121}$$

and write the equation of motion for the x-direction:

$$a_x = g\sin(\theta) \tag{1.122}$$

which is a constant.

From this point on the solution should be familiar – since $v_y(0) = 0$ and $y(0) = 0$, $y(t) = 0$ and we can ignore y altogether and the problem is now one dimensional! See if you can find how long it takes for the block to reach bottom, and how fast it is going when it gets there. You should find that $v_{\mathrm{bottom}} = \sqrt{2gH}$, a familiar result (see the very first example of the dropped ball) that suggests that there is more to learn, that gravity is somehow "special" if a ball can be dropped *or slide* down from a height H and reach the bottom going at the same speed either way!

1.9: Circular Motion

So far, we've solved only two dimensional problems that involved a constant acceleration in some specific direction. Another very general (and important!) class of motion is *circular motion*. This could be: a ball being whirled around on a string, a car rounding a circular curve,

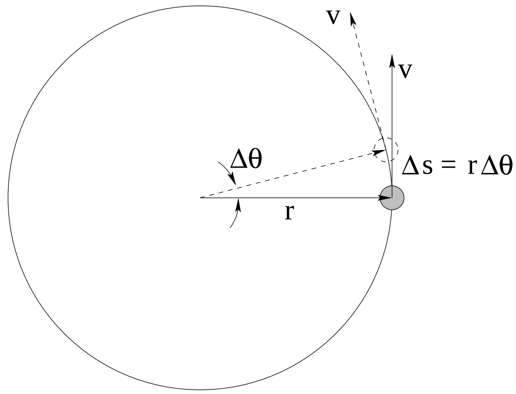


Figure 1.18: A way to visualize the motion of a particle, e.g. a small ball, moving in a circle of radius r. We are looking down from above the circle of motion at a particle moving counterclockwise around the circle. At the moment, at least, the particle is moving at a constant speed v (so that its *velocity* is always *tangent* to the circle.

a roller coaster looping-the-loop, a bicycle wheel going round and round, almost *anything* rotating about a fixed axis has all of the little chunks of mass that make it up going in circles!

Circular motion, as we shall see, is "special" because the acceleration of a particle moving in a circle towards the center of the circle has a value that is completely determined by the *geometry* of this motion. The form of centripetal acceleration we are about to develop is thus a *kinematic* relation – not dynamical. It doesn't matter *which* force(s) or force rule(s) off of the list above make something actually move around in a circle, the relation is true for all of them. Let's try to understand this.

1.9.1: Tangential Velocity

First, we have to visualize the motion clearly. Figure 1.18 allows us to see and think about the motion of a particle moving in a circle of radius r (at a constant *speed*, although later we can relax this to *instantaneous* speed) by visualizing its *position* at two successive times. The first position (where the particle is solid/shaded) we can imagine as occurring at time t. The second position (empty/dashed) might be the position of the particle a short time later at (say) $t + \Delta t$.

During this time, the particle travels a short distance around the arc of the circle. Because *the length of a circular arc is the radius times the angle subtended by the arc* we can see that:

$$\Delta s = r\Delta\theta \tag{1.123}$$

Note Well! In this and all similar equations θ must be measured in **radians**, never degrees. In fact, angles measured in degrees are fundamentally meaningless, as degrees are an arbitrary partitioning of the circle. Also note that radians (or degrees, for that matter) are *dimensionless* – they are the *ratio* between the length of an arc and the radius of the arc (think 2π is the ratio of the circumference of a circle to its radius, for example).

The *average speed* v of the particle is thus this distance divided by the time it took to move it:

$$v_{\rm avg} = \frac{\Delta s}{\Delta t} = r\frac{\Delta\theta}{\Delta t} \quad (1.124)$$

Of course, we really don't want to use average speed (at least for very long) because the speed might be varying, so we take the limit that $\Delta t \to 0$ and turn everything into derivatives, but it is much easier to draw the pictures and *visualize* what is going on for a small, finite Δt:

$$v = \lim_{\Delta t\to 0} r\frac{\Delta\theta}{\Delta t} = r\frac{d\theta}{dt} \quad (1.125)$$

This speed is directed *tangent to the circle of motion* (as one can see in the figure) and we will often refer to it as the ***tangential velocity***. Sometimes I'll even put a little "t" subscript on it to emphasize the point, as in:

$$v_t = r\frac{d\theta}{dt} \quad (1.126)$$

but since the velocity is *always* tangent to the trajectory (which just happens to be circular in this case) we don't really need it.

In this equation, we see that the speed of the particle at any instant is the radius times the rate that the angle is being swept out by by the particle per unit time. This latter quantity is a very, very useful one for describing circular motion, or rotating systems in general. We define it to be the ***angular velocity***:

$$\Omega = \frac{d\theta}{dt} \quad (1.127)$$

Thus:

$$v = r\Omega \quad (1.128)$$

or

$$\Omega = \frac{v}{r} \quad (1.129)$$

are both extremely useful expressions describing the kinematics of circular motion.

1.9.2: A Note on Notation

In the previous section we used the symbol "capital omega" – Ω – to stand for *angular velocity*. If you compare this textbook with many, if not most, other introductory physics textbooks, you will observe that it is very common to use "lower-case omega" – ω – for this, that is:

$$\omega = \frac{d\theta}{dt} = \frac{v}{r}$$

for a particle (or later, chunk of mass dm) moving in a circle of radius r.

There is one problem with this. Eventually we will study ***simple harmonic oscillation***, and two of the oscillators we will look at are *pendulums* of a variety of shapes and arrangements, and *torsional oscillators*. In both cases, the system has *both* angular *velocity* – the pendulum moves along a circular arc, the torsional oscillator rotates around the axis of a torsional spring – *and* something we will define in that chapter called *angular frequency*, and those same textbooks invariably use the same symbol, ω, for both of these very different quantities. A student may find themselves writing as part of their answer to the question: "what is the angular velocity of the swinging pendulum bob?" something like

$$\omega = -\omega\theta_0 \sin(\omega t + \phi)$$

and trying to differentiate the two quantities with an ambiguous subscript such as:

$$\omega_a = -\omega\theta_0 \sin(\omega t + \phi)$$

where in this expression:

$$\omega_a = \frac{d\theta}{dt} \qquad \text{(angular } \textit{velocity}\text{)}$$

and

$$\omega = \sqrt{\frac{g}{\ell}} \qquad \text{(angular } \textit{frequency}\text{)}$$

The two are differentiated by the fact that harmonic oscillators oscillate with a harmonic angular frequency even when absolutely nothing in the system actually rotates through an angle, while angular velocity is simply the rate at which *something* sweeps out an angle as it moves relative to some selected pivot/axis. Masses on springs have angular frequencies. Linearly polarized light waves have angular frequencies. Quantum mechanical wavefunctions have angular frequencies. In none of these cases is anything at all sweeping out a physical angle or rotating.

This can seriously disturb introductory students, who are not yet practiced at using the same symbols for different physical quantities in *different* contexts, let alone using the same symbols for different physical quantities in a single problem and context! Nor are they yet prepared for the idea that one sometimes might write (for example) the equation for the x-component of a particle moving in a circle of radius r at constant speed $x(t) = r\cos(\Omega t)$ where Ω is the angular velocity, but its use in the context of this expression for the x component only is more as an angular *frequency*.

For that reason I have elected to differentiate the symbols explicitly in this book, so that the offending equation for angular velocity would be written as:

$$\Omega = -\omega\theta_0 \sin(\omega t + \phi)$$

without the slightest ambiguity. This is the angular velocity of the pendulum; the ω is its angular frequency.

Be alert for a few other discussions of notational differences in other places in this textbook. It is commonplace to use λ, for example, as both a symbol for linear mass density and for wavelength of a wave. When studying waves on strings, however, both again occur in the same context, and can actually occur with both meanings in a single equation! Again expert textbook writers and second or third year physics majors may not even notice this, or may throw in a subscript to differentiate them notationally, but undergraduates will simply become confused and sad and make entirely unnecessary mistakes in their algebra or due to confused conceptual thinking that after all is not really their fault. For that reason, we will try to use symbols that collide a bit *less* whenever we can do so without departing too much from the norm – ℓ for lengths instead of L as the latter is also the universal symbol for angular momentum, and rods of length L may well have an angular momentum L that depends on L as well (see how difficult this can be?); μ for linear mass density instead of λ, as it is less likely to find linear mass density and dynamic viscosity or magnetic moment (two other common uses for the symbol μ) in the same problem.

1.9.3: Centripetal Acceleration

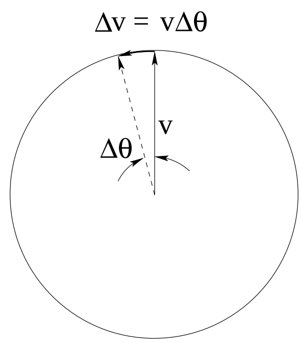


Figure 1.19: The velocity of the particle at t and $t + \Delta t$. Note that over a very short time Δt the speed of the particle is at least approximately constant, but its *direction* varies because it always has to be perpendicular to $\vec{r}$, the vector from the center of the circle to the particle. The velocity swings through the *same angle* $\Delta\theta$ that the particle itself swings through in this (short) time.

Next, we need to think about the velocity of the particle (not just its speed, note well, we have to think about direction). In figure 1.19 you can see the velocities from figure 1.18 at time t and $t + \Delta t$ placed so that they begin at a common origin (remember, you can move a vector anywhere you like as long as the magnitude and direction are preserved).

The velocity is perpendicular to the vector $\vec{r}$ from the origin to the particle at any instant of time. As the particle rotates through an angle $\Delta\theta$, the velocity of the particle *also* must rotate through the angle $\Delta\theta$ while its magnitude remains (approximately) the same.

In time Δt, then, the magnitude of the *change* in the velocity is:

$$\Delta v = v\Delta\theta \tag{1.130}$$

Consequently, the average magnitude of the acceleration is:

$$a_{\rm avg} = \frac{\Delta v}{\Delta t} = v\frac{\Delta\theta}{\Delta t} \tag{1.131}$$

As before, we are interested in the instantaneous value of the acceleration, and we'd also like to determine its *direction* as it is a vector quantity. We therefore take the limit $\Delta t \to 0$ and inspect the figure above to note that the direction in that limit is to the left, that is to say *in the negative* $\vec{r}$ *direction!* (You'll need to look at both figures, the one representing position and the other representing the velocity, in order to be able to see and understand this.) The instantaneous magnitude of the acceleration is thus:

$$a = \lim_{\Delta t\to 0} v\frac{\Delta\theta}{\Delta t} = v\frac{d\theta}{dt} = v\Omega = \frac{v^2}{r} = r\Omega^2 \tag{1.132}$$

where we have substituted equation 1.129 for Ω (with a bit of algebra) to get the last couple of equivalent forms. The direction of this vector is *towards the center of the circle.*

The word "centripetal" means "towards the center", so we call this *kinematic* acceleration the ***centripetal acceleration*** of a particle moving in a circle and will often label it:

$$a_c = v\Omega = \frac{v^2}{r} = r\Omega^2 \tag{1.133}$$

A second way you might see this written or referred to is as the r-component of a vector in *plane polar coordinates*. In that case "towards the center" is in the $-\hat{\boldsymbol{r}}$ direction and we could write:

$$a_r = -v\Omega = -\frac{v^2}{r} = -r\Omega^2 \tag{1.134}$$

In most actual problems, though, it is easiest to just compute the magnitude a_c and then assign the direction in terms of the particular coordinate frame *you* have chosen for the problem, which might well make "towards the center" be the positive x direction or something else entirely in *your* figure at the instant drawn.

This is an *enormously useful result*. Note well that it is a *kinematic* result – math with units – not a *dynamic* result. That is, I've made no reference whatsoever to *forces* in the derivations above; the result is a pure mathematical consequence of motion in a 2 dimensional plane circle, quite independent of the particular forces that *cause* that motion. The way to think of it is as follows:

> ***If*** a particle is moving in a circle at instantaneous speed v, ***then*** its acceleration towards the center of that circle is v^2/r (or $r\Omega^2$ if that is easier to use in a given problem).

This specifies the *acceleration* in the component of Newton's Second Law that points towards the center of the circle of motion! No matter *what* forces act on the particle, if it moves in a circle the component of the *total* force acting on it towards the center of the circle must be $ma_c = mv^2/r$. If the particle is moving in a circle, then the centripetal component of the total force must have this value, but this quantity isn't itself a force law or rule! *There is no such thing as a "centripetal force"*, although there are *many* forces that can cause a centripetal acceleration in a particle moving in circular trajectory.

Let me say it again, with emphasis: A common mistake made by students is to confuse mv^2/r with a "force rule" or "law of nature". It is **nothing of the sort**. No special/new force "appears" because of circular motion, the circular motion is caused by the usual forces we list above in some combination that *add up* to $ma_c = mv^2/r$ in the appropriate direction. Don't make this mistake one a homework problem, quiz or exam! Think about this a bit and discuss it with your instructor if it isn't completely clear.

Example 1.9.1: Ball on a String

We wish to find the tension T in a string as a ball of mass m swings down in a circular arc of radius L. In a later chapter, we'll learn to use energy conservation to find its speed v at the

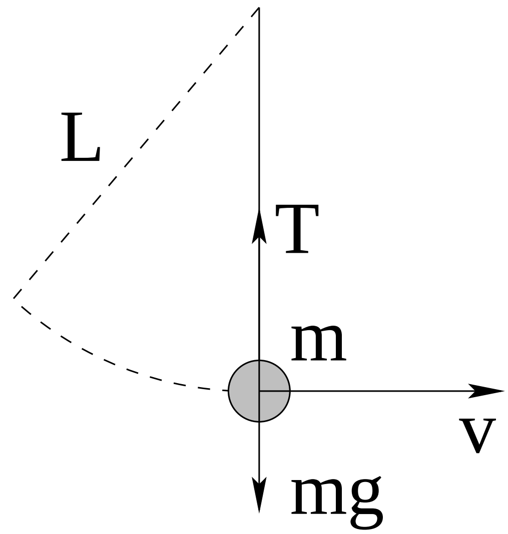


Figure 1.20: A ball of mass m swings down in a circular arc of radius L suspended by a string, arriving at the bottom with speed v. What is the tension in the string?

bottom of its arc given information about how it starts, but for now we'll treat its speed v there as given and we will neglect drag. This situation is portrayed in figure 1.20.

Observe that at the bottom of the trajectory, the tension T in the string points *straight up* (towards the center of the circle of motion) and the force mg points (conveniently enough) straight *down*. No other forces act, so we should choose coordinates such that one axis lines up with these two forces. Let's use $+y$ vertically up, aligned with the string (and note that at this instant, there are no forces acting in the x direction. Then:

$$F_y = T - mg = ma_y = m\frac{v^2}{L} \tag{1.135}$$

or

$$T = mg + m\frac{v^2}{L} \tag{1.136}$$

Wow, that was easy! Easy or not, this simple example is a *very useful one* as it will form part of the solution to *many* of the problems you will solve in the next few weeks, so be sure that you understand it. The *net* force towards the center of the circle must be algebraically equal to mv^2/r, where I've cleverly given you L as the radius of the circle instead of r just to see if you're paying attention[63].

Example 1.9.2: Tether Ball/Conic Pendulum

Suppose you hit a tether ball (a ball on a string or rope, also called a conic pendulum as the rope sweeps out a right circular cone) so that it moves in a plane circle at an angle θ at the end of a string of length L. Find T (the tension in the string) and v, the speed of the ball such that this is true.

[63]There is actually an important lesson here as well: *Read the problem!* I can't tell you how often students miss points because they don't solve the problem given, they solve a problem *like* the problem given that perhaps was a class example or on their homework. This is easily avoided by reading the problem carefully and using the variables and quantities *it* defines. Read the problem!

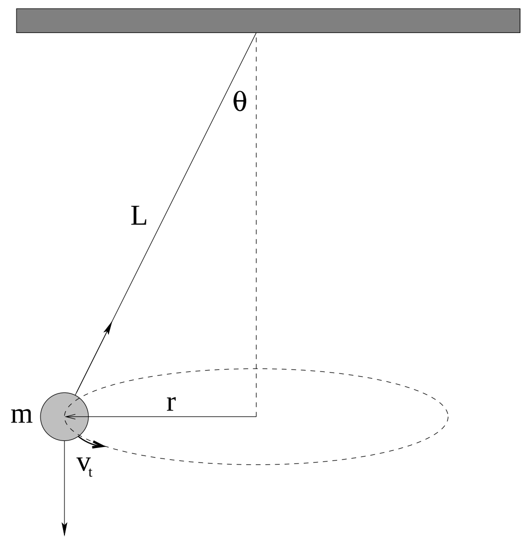


Figure 1.21: Ball on a rope (a tether ball or conical pendulum). The ball sweeps out a right circular cone at an angle θ with the vertical when launched appropriately.

We note that if the ball is moving in a circle of radius $r = L\sin\theta$, its centripetal acceleration *must* be $a_r = -\frac{v^2}{r}$. Since the ball is not moving up and down, the vertical forces must cancel. This suggests that we should use a coordinate system with $+y$ vertically up and x in towards the center of the circle of motion, *but* we should bear in mind that we will also be thinking of the motion in plane polar coordinates *in* the plane and that the angle θ is specified relative to the vertical! Oooo, head aching, must remain calm and visualize, visualize.

Visualization is aided by a good figure, like the one (without coordinates, you can add them) in figure 1.21. Note well in this figure that the only "real" forces acting on the ball are gravity and the tension T in the string. Thus in the y-direction we have:

$$\sum F_y = T\cos\theta - mg = 0 \tag{1.137}$$

and in the x-direction (the minus r-direction, as drawn) we have:

$$\sum F_x = T\sin\theta = ma_r = \frac{mv^2}{r}. \tag{1.138}$$

Thus

$$T = \frac{mg}{\cos\theta}, \tag{1.139}$$

$$v^2 = \frac{Tr\sin\theta}{m} \tag{1.140}$$

Eliminating T and substituting $r = L\sin\theta$ to get the answer in terms of the givens, we get:

$$v = \sqrt{gL\sin\theta\tan\theta} \tag{1.141}$$

Nobody said all of the answers will be pretty... but this is still pretty easy to evaluate.

1.9.4: Tangential Acceleration

Sometimes we will want to solve problems where a particle speeds up or slows down *while* moving in a circle. Obviously, this means that there is a nonzero *tangential acceleration* changing the *magnitude* of the tangential velocity.

Let's write $\vec{F}$ (total) acting on a particle moving in a circle in a coordinate system that rotates along with the particle – plane polar coordinates. The tangential direction is the $\hat{\theta}$ direction, so we will get:

$$\vec{F} = F_r\hat{r} + F_t\hat{\theta} \tag{1.142}$$

From this we will get two equations of motion (connecting this, at long last, to the dynamics of two dimensional motion):

$$F_r = -m\frac{v^2}{r} \tag{1.143}$$

$$F_t = ma_t = m\frac{dv}{dt} \tag{1.144}$$

The acceleration on the *right* hand side of the first equation is determined from m, v, and r, but $v(t)$ itself is determined from the *second* equation. You will use these two equations *together* to solve the "bead sliding on a wire" problem in the next week's homework assignment, so keep this in mind.

That's about it for the first week. We have more to do, but to do it we'll need more forces. Next week we move on to learn some more forces from our list, especially *friction* and *drag forces*. We'll wrap the week's work up with a restatement of our solution rubric for "standard" dynamics problems. I would recommend literally ticking off the steps in your mind (and maybe on the paper!) as you work this week's homework. It will really help you later on!

1.10: Conclusion: Rubric for Newton's Second Law Problems

a) *Draw* a good picture of what is going on. In general you should probably do this even if one has been provided for you – visualization is key to success in physics.

b) On your drawing (or on a second one) decorate the objects with all of the *forces* that act on them, creating a *free body diagram* for the forces on each object.

c) *Write* Newton's Second Law for each object (summing the forces and setting the result to $m_i\vec{a}_i$ for each – ith – object) and algebraically rearrange it into (vector) differential equations of motion (practically speaking, this means solving for or isolating the *acceleration* $\vec{a}_i = \frac{d^2\vec{x}_i}{dt^2}$ of the particles in the equations of motion).

d) *Decompose* the 1, 2 or 3 dimensional equations of motion for each object into a *set of independent* 1 dimensional equations of motion for each of the orthogonal coordinates by choosing a suitable coordinate system (which may *not* be cartesian, for some problems) and using trig/geometry. Recall the rule above, and try to pick coordinates where one or more axes are in directions where we know the acceleration from the problem constraints, for example directions where it is zero or v^2/r, so that just one axis points in a direction where you have to use Newton's Second Law to actually solve for nontrivial motion.

ote that a "coordinate" here may even wrap around a corner following a string, for example – or we can use a different coordinate system for each particle, as long as we ***have a known relation between the coordinate systems***. And *use* it to ultimately answer the questions!

e) *Solve* the independent 1 dimensional systems for each of the independent orthogonal coordinates chosen, plus any coordinate system constraints or relations. In many problems the *constraints* will eliminate one or more degrees of freedom from consideration (if we've chosen our coordinates wisely, for example). Note that in most nontrivial cases, these solutions will have to be *simultaneous* solutions, obtained by e.g. algebraic substitution or elimination.

f) *Reconstruct* the multidimensional trajectory by adding the vectors components thus obtained back up (for a common independent variable, time).

g) *Answer* algebraically any questions requested concerning the resultant trajectory.

Homework for Week 1

Before you Begin...

There are "no numbers" in most of the homework problems in this text. **This is deliberate** – algebra is a ***reasoning*** tool and physics is all about empirically founded reason! Arithmetical evaluation of formulas given numerical data on their contents, on the other hand is a process of more or less mechanical substitution and evaluation – often irreverently referred to as "plug and chug" – that can be and often is performed by entities that understand nothing at all about the origins or meaning of the formula into which numbers are being substituted or the result of the computation[64].

That is not meant to suggest that arithmetical practice is useless in physics problem, only to explain why this text de-emphasizes it. Arithmetic's primary virtue in physics *practice* problems is to permit students to get a concrete feel for reasonable/typical sizes or scales of real-work results *once a student understands those results!* A secondary virtue is that well, yeah, physics *is* supposedly a quantitatively precise theory of how everything works and one needs numbers in order to compare that theory with reality via measured experimental numbers – the basis for the lab part of a typical physics course. On both grounds, a physicist should never be completely *incompetent* at arithmetic even when done by hand, and the ability to perform quick and accurate *numerical estimates* of results has long been prized.

In each of the following chapters, most of the provided homework problems are intended for all students of physics and should *all* be completed by those students at the end of each week/chapter. They are sometimes followed by a few clearly marked "advanced" problems that are intended to be assigned primarily to physics majors, math majors taking physics, or engineering students, who are expected to know and be able to skillfully use a bit more general mathematics (especially calculus) than e.g. life science students, but note well that there is *plenty of math including calculus in the general problems* even for non-major life science students. It is impossible to learn and understand physics without at least *some* competence in calculus[65]. Newton invented calculus *just so he could formulate physics* and this course *teaches and reinforces* the use of algebra, geometry, trigonometry and calculus both to permit all of classical physics to be *consistently developed* from Newton's Laws and a handful of empirical (e.g. force) laws and to solve problems that exemplify and illuminate each new set of concepts and results as they are developed.

Please do not skimp on or skip the homework, if you are using this text to learn physics! Students who work homework problems to ***mastery*** – the state where you can do each assigned problem, perfectly, ***without using any external resource including your notes or the textbook itself*** – will almost certainly excel in the course and earn high marks as a result. Students that only work hard enough to *barely get through the homework* with the book in one hand and their lecture notes in the other or worse, don't *even* honestly complete the homework via personal struggle and effort but copy the work of others or hand it in incomplete, well, what does your *reason* tell you is a likely outcome gradewise when confronted with problems you still haven't mastered and don't really understand on a test?

---

[64]Such as computers, or students armed with calculators who were taught physics as a pile of formulas to be memorized instead of understood.

[65]It is the opinion of the author that so-called "algebraic physics" is taught as an empty exercise in the memorization of formulas whose origins are concealed from the student, shrouded in the mists of *calculus*...

Problem 1.

Physics Concepts

In order to solve the following physics problems for homework, you will need to have the following physics and math concepts first at hand, then in your long term memory, ready to bring to bear whenever they are needed. Every week (or day, in a summer course) there will be new ones.

To get them there efficiently, you will need to carefully organize what you learn as you go along. This organized summary will be a *standard, graded part of every homework assignment!*

Your homework will be graded in two *equal* parts. Ten points will be given for a complete crossreferenced summary of the physics concepts used in each of the assigned problems. One problem will be selected for grading in detail – usually one that well-exemplifies the material covered that week – for ten more points.

Points will be taken off for egregiously missing concepts or omitted problems in the concept summary. Don't just name the concepts; if there is an equation and/or diagram associated with the concept, put that down too. Indicate (by number) all of the homework problems where a concept was used.

This concept summary will eventually help you prioritize your study and review for exams! To help you understand what I have in mind, I'm building you a list of the concepts for *this* week, and indicating the problems that (will) need them as a sort of template, or example. However, ***Note Well!*** You must write up, and hand in, your *own* version ***this week*** as well as all of the other weeks to get full credit.

In the end, if you put your homework assignments including the summaries for each week into a three-ring binder as you get them back, you will have a nearly perfect ***study guide*** to go over before all of the exams and the final. You might want to throw the quizzes and hour exams in as well, as you get them back. Remember the immortal words of Edmund Burke: "Those who don't know history are destined to repeat it" – know your own "history", by carefully saving, and going over, your own work throughout this course!

- Writing a vector in cartesian coordinates. For example:

$$\vec{A} = A_x \hat{x} + A_y \hat{y} + A_z \hat{z}$$

  Used in problems 2,3,4,5,6,7,8,9,10,12

- Decomposition of a vector at some angle into components in a (2D) coordinate system. Given a vector $\vec{A}$ with length A at angle θ with respect to the x-axis:

$$A_x = A\cos(\theta)$$

$$A_y = A\sin(\theta)$$

  Used in problem 5,6,9,10,11,12

- Definition of trajectory, velocity and acceleration of a particle:

  The trajectory is the vector $\vec{x}(t)$, the vector *position* of the particle as a function of the *time*.

  The velocity of the particle is the (vector) rate at which its position changes as a function of time, or the time derivative of the trajectory:

$$\vec{v} = \frac{\Delta \vec{x}}{\Delta t} = \frac{d\vec{x}}{dt}$$

  The acceleration is the (vector) rate at which its velocity changes as a function of time, or the time derivative of the velocity:

$$\vec{a} = \frac{\Delta \vec{v}}{\Delta t} = \frac{d\vec{v}}{dt}$$

  Used in all problems.

- Inertial reference frame

  A set of coordinates in which (if you like) the laws of physics that describe the trajectory of particles take their simplest form. In particular a frame in which Newton's Laws (given below) hold in a consistent manner. A set of coordinates that is not itself accelerating with respect to all of the other non-accelerating coordinate frames in which Newton's Laws hold.

  Used in all problems (when I choose a coordinate system that is an inertial reference frame).

- Newton's First Law

  In an inertial reference frame, an object in motion will remain in motion, and an object at rest will remain at rest, unless acted on by a net force.

  If $\vec{F} = 0$, then $\vec{v}$ is a constant vector.

  A consequence, as one can see, of Newton's Second Law. Not used much yet.

- Newton's Second Law

  In an inertial reference frame, the net vector force on an object equals its mass times its acceleration.

$$\vec{F} = m\vec{a}$$

  Used in every problem! **Very important!** Key! Five stars! *****

- Newton's Third Law

  If one object exerts a force on a second object (along the line connecting the two objects), the second object exerts an equal and opposite force on the first.

$$\vec{F}_{ij} = -\vec{F}_{ji}$$

  Not used much yet.

- Differentiating x^n

  $$\frac{dx^n}{dx} = nx^{n-1}$$

  Not used much yet.

- Integrating $x^n dx$

  $$\int x^n dx = \frac{x^{n+1}}{n+1}$$

  Used in every problem where we implicitly use kinematic solutions to constant acceleration to find a trajectory.

  Problems 2

- The force exerted by gravity near the Earth's surface

  $$\vec{\boldsymbol{F}} = -mg\hat{\boldsymbol{y}}$$

  (down).

  Used in problems 2,3,4,5,6,8,9,10,11,12

  Problems 2

- Centripetal acceleration.

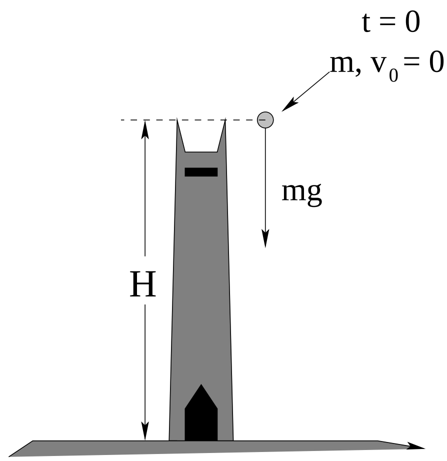
  $$a_r = -\frac{v^2}{r}$$

  Used in problems 11,13

This isn't a *perfect* example – if I were doing this by hand I would have drawn pictures to accompany, for example, Newton's second and third law, the circular motion acceleration, and so on.

I also included more concepts than are strictly needed by the problems – *don't hesitate* to add important concepts to your list even if none of the problems seem to need them! Some concepts (like that of inertial reference frames) are *ideas* and underlie problems even when they aren't actually/obviously used in an algebraic way in the solution!

Problem 2.



A ball of mass m is dropped at time $t = 0$ from the top of the Duke Chapel (which has height H) to fall freely under the influence of *gravity*. Neglect "wind resistance" (which we'll come to call drag in the next chapter).

a) How long does it take for the ball to reach the ground?

b) How fast is it going when it reaches the ground?

Just this once, for this *first* problem of many, many you will solve as the course proceeds, I'm going to give you a bunch of advice on how to get started, but remember to look back at the rubric for solving force problems in the chapter content as well as the following more or less recapitulates this rubric. To solve this first problem:

- Draw a *good figure* – in this case a chapel tower, the ground, the ball falling. Label the distance H in the figure, indicate the force on the mass with a vector arrow labelled mg pointing *down*. This is called a *force diagram*. Alternative, draw an "inset" figure of the mass(es) off to the side and decorate it *alone* with the forces acting on it (maintaining their coordinate orientation). This is called a *free body diagram* as it concentrates on each body separately, "free" from the others. Note well! Solutions without a figure (usually including either a force diagram or free body diagram) will lose points!

- *Choose coordinates!* In this case you could (for example) put an origin at the bottom of the tower with a y-axis going up so that the height of the object is $y(t)$.

- Write Newton's second law for the mass.

- Transform it into a (differential) equation of motion. This is the *math* problem that must be solved.

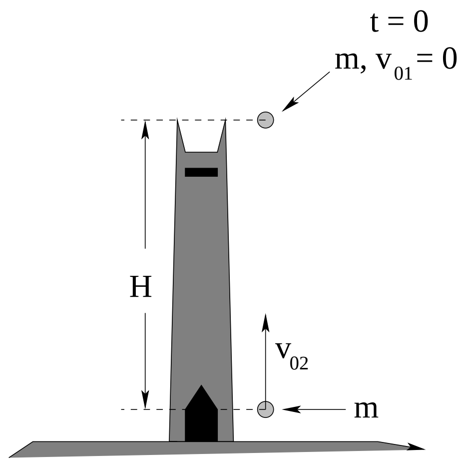
- In this case, you will want to integrate the *constant* $\frac{dv_y}{dt} = a_y = -g$ to get $v_y(t)$, then integrate $\frac{dy}{dt} = v_y(t)$ to get $y(t)$.

- Express the algebraic condition that is true when the mass reaches the ground, and solve for the *time* it does so, answering the first question.

- Use the answer to the first question (plus your solutions) to answer the second. These last two steps requires a mix of ***creative thinking*** and ***experience*** to give you the *insight* as to how to proceed, and the following assignments will ensure that you get a *lot* of practice at this so that you become quite good at it!

The first four steps in this solution will nearly always be the same for Newton's Law problems (and, with small modifications, for most of the problems in this textbook!). Once one has the equation of motion, solving the *rest* of the problem depends on the force law(s) in question, and answering the questions requires a bit of insight that only comes from practice, practice, practice. So practice!

Problem 3.



A baseball of mass m is dropped at time $t = 0$ from rest ($v_{01} = 0$) from the top of the Duke Chapel (which has height H) to fall freely under the influence of *gravity*. At the same instant, a *second* baseball of mass m is thrown *up* from the ground directly beneath at a speed v_{02} (so that if the two balls travel far enough, fast enough, they will collide). In answering the following questions, neglect drag.

a) Draw a force diagram or free body diagram for each mass, and then compute the net force acting on each mass, ***separately***. You can neglect all directions but the vertical direction (so this is a "one dimensional (1D) motion problem").

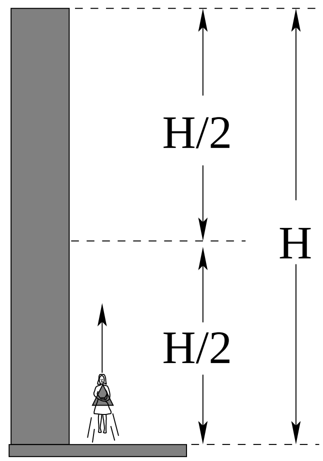
b) ***From the equation of motion for each mass***, determine their one dimensional trajectory functions, $y_1(t)$ and $y_2(t)$. Please actually do the two integrals (and use the initial conditions) for each mass, don't just look up or remember the solution.

c) Sketch a *qualitatively correct* graph of $y_1(t)$ and $y_2(t)$ on the same set of axes in the case where the two collide before they hit the ground, and draw a *second* graph of $y_1(t)$ and $y_2(t)$ on a new set of axes in the case where they do not. From your two pictures, determine a *criterion* for whether or not the two balls will actually collide before they hit the ground. Express this criterion as an algebraic expression (inequality) involving H, g, and v_{02}.

d) The Duke Chapel is roughly 100 meters high. What (also roughly, you may estimate and don't need a calculator) is the minimum velocity v_{02} a the second mass must be thrown up in order for the two to collide? Note that you should give an actual numerical answer here. What is the (again approximate, no calculators) answer in miles per hour, assuming that 1 meter/second $\approx$ 9/4 miles per hour? Do you think you can throw a baseball that fast?

Pro tip: The $y(t)$ functions for both masses will turn out to be *upside down parabolas* with *exactly the same curvature*. They differ only in their y-intercept y_0 at $t = 0$ and their slope v_0 at $t = 0$. If you want to make a *neat* diagram, draw a single, very neat parabola, turn it upside down and then trace it (with suitable intercepts) onto a single y-t frame for various possible intercepts for the lower ball. Can you understand the question now graphically?

Problem 4.



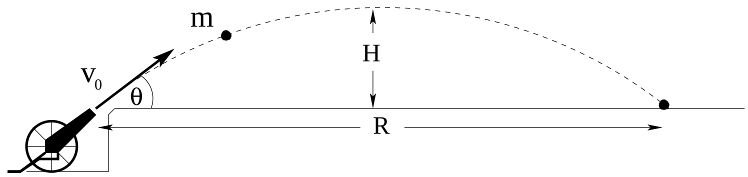
Supergirl launches herself into the sky to fly to the top of a tall building of height H to rescue a cat. Using her Kryptonian powers, she accelerates upward with an acceleration of $g/3$. After a time t_1 she has reached the height $H/2$. In an ***additional*** time t_2, she reaches the top of the building (so she reaches its top in a total time $t_t = t_1 + t_2$) and comes "instantly" to rest.

a) What is the height H of the building in terms of the time t_2?

b) The cat turned out to be a supercat and flew off in the meantime, so she jumps back to the ground, allowing herself to fall freely. How long t_3 does it take for her to reach the ground?

c) If $t_2 = 2$ seconds, what is H in meters and t_3 in seconds. Use $g \approx 10$ meters/second^2, and be careful about significant figures!

Problem 5.



A cannon sits on a horizontal plain. It fires a cannonball of mass m at speed v_0 at an angle θ relative to the ground. With a mix of algebra, trig, and calculus find:

a) The maximum height H of the cannonball's trajectory.

b) The time t_a the cannonball is in the air.

c) The range R of the cannonball.

d) Use the trigonometric identity:

$$2\sin(\theta)\cos(\theta) = \sin(2\theta)$$

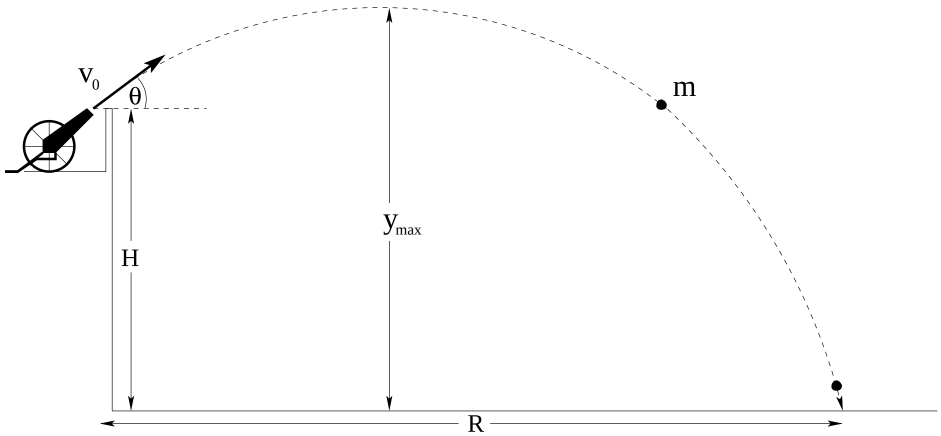
to express your result for the range. For a fixed v_0, how many angles (usually) can you set the cannon to that will have the same range?

Some *qualitative* or *conceptual* questions follow. You may not *need* to use your algebraic answers per se to answer conceptual questions once you understand things like scaling or "how things work"!

e) How does the time the cannonball remains in the air depend on its maximum height?

f) If the cannon is fired at several different angles and initial speeds, does the combination of angle and speed with the greatest ***range*** always remain in the air the longest?

Problem 6.



A cannon sits on at the top of a rampart of height (to the mouth of the cannon) H. It fires a cannonball of mass m at speed v_0 at an angle θ relative to the ground. Find:

a) The maximum height $y_{\rm max}$ of the cannonball's trajectory.

b) The time the cannonball is in the air.

c) The range of the cannonball.

A conceptual question:

d) In your solution to b) above you should have found *two* times that the height of the cannonball is "ground level", one of them ***negative***. What does the negative time correspond to? (Consider: does your mathematical solution "know" about the *actual* history of the cannonball before you fired the cannon?)

You might find the quadratic formula useful in solving this problem. We will be using this *a lot* in this course, and on a quiz or exam you won't be given it, so be sure that you *really* learn it now in case you don't know or have forgotten it. The roots of a quadratic:

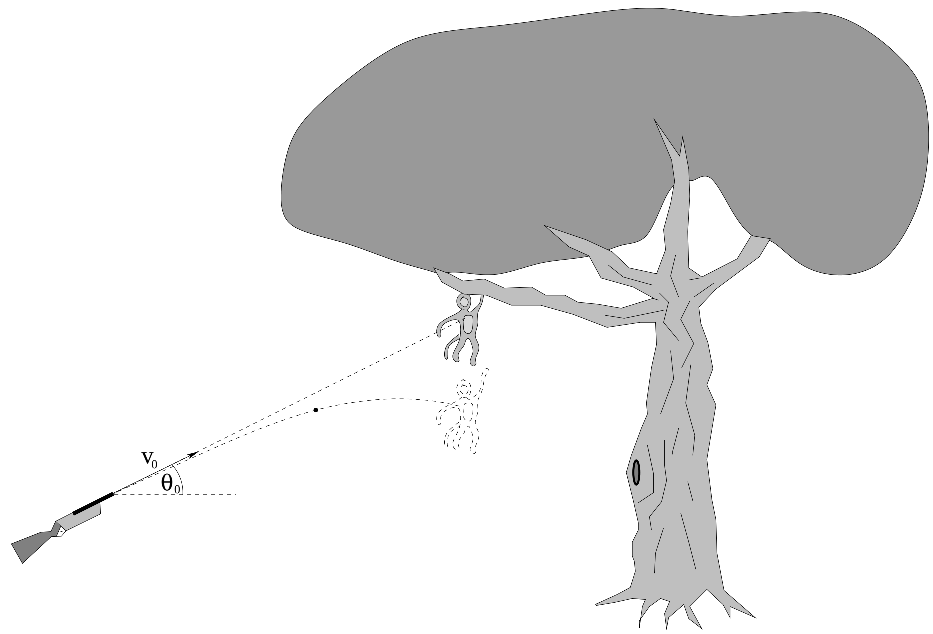
$$ax^2 + bx + c = 0$$

are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

You can actually derive this for yourself if you like (often, mastering a derivation of things helps you remember them). Just i) divide the original quadratic by a; ii) complete the square by adding and subtracting the 'right' algebraic quantity; and iii) then evaluate the square root of the perfect square. Simple!

Problem 7.



A researcher aims her tranquiler gun (that fires a dart at initial speed v_0; ignore drag) at a monkey in a distant tree. Just as she fires, the monkey lets go and drops in free fall towards the ground.

a) Show that the sleeping dart hits the monkey independent of v_0 as long as the angle θ_0 points *directly* at the monkey!

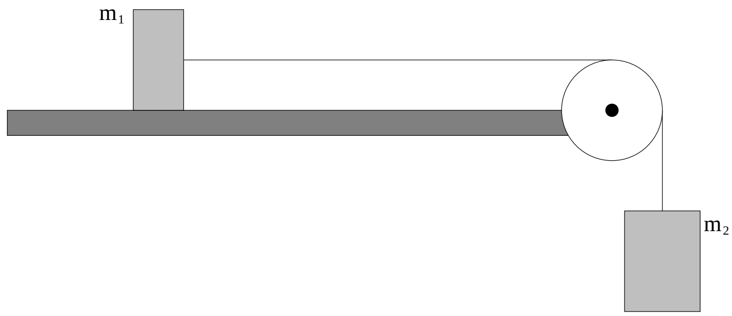
There are some unstated assumptions in this problem in order for your answer to actually make sense. Here are some conceptual questions to help you understand this:

b) If the gun shoots the dart too slowly (v_0 "too small"), what will *really* happen to the dart? In particular, will it hit the monkey, and if not, what will it actually hit?

Of course, real guns usually fire a bullet that moves so fast that the trajectory at any reasonable range is quite flat. Real hunters adjust their sights to allow for the drop in their dart/bullet, basically aiming the gun at a point *above* the monkey so it hits it if the monkey does *not* let go and drop the instant the gun is fired. And then there *is* drag, covered in the next chapter!

Be at peace. No monkeys, real or virtual, were harmed in this problem.

Problem 8.



A mass m_1 is attached to a second mass m_2 by an Acme (massless, unstretchable) string. m_1 sits on a frictionless table; m_2 is hanging over the ends of a table, suspended by the taut string from an Acme (frictionless, massless) pulley. At time $t = 0$ both masses are released.

a) Draw the force/free body diagram for this problem.

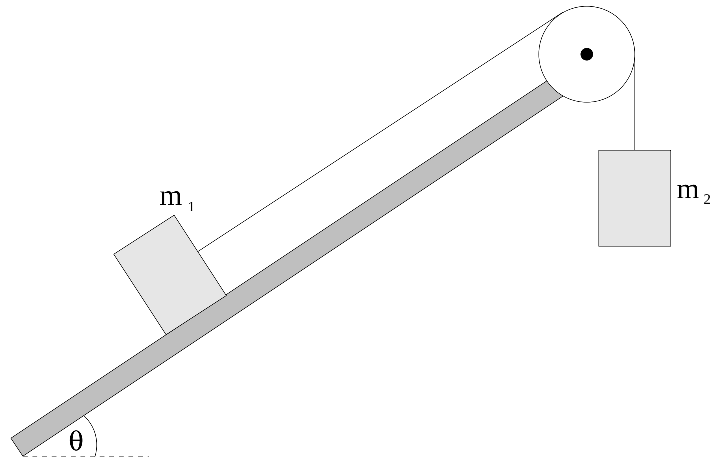
b) Find the acceleration of the two masses.

c) The tension T in the string.

d) How fast are the two blocks moving when mass m_2 has fallen a height H (assuming that m_1 hasn't yet hit the pulley)?

Conceptual discussion: Your answer to b) should look something like: The *total unopposed external force* (gravity acting on m_2) seems to accelerate *both* masses 'as if they were one'. The string just *transfers* force from one mass to the other so that they accelerate *together!* This is a common feature to many problems involving multiple masses and a mix of internal forces, as we'll see and eventually derive and formalize!

Problem 9.



A mass m_1 is attached to a second mass $m_2 > m_1$ by an Acme (massless, unstretchable) string. m_1 sits on a frictionless inclined plane at an angle θ with the horizontal; m_2 is hanging over the high end of the plane, suspended by the taut string from an Acme (frictionless, massless) pulley. At time $t = 0$ both masses are released from rest.

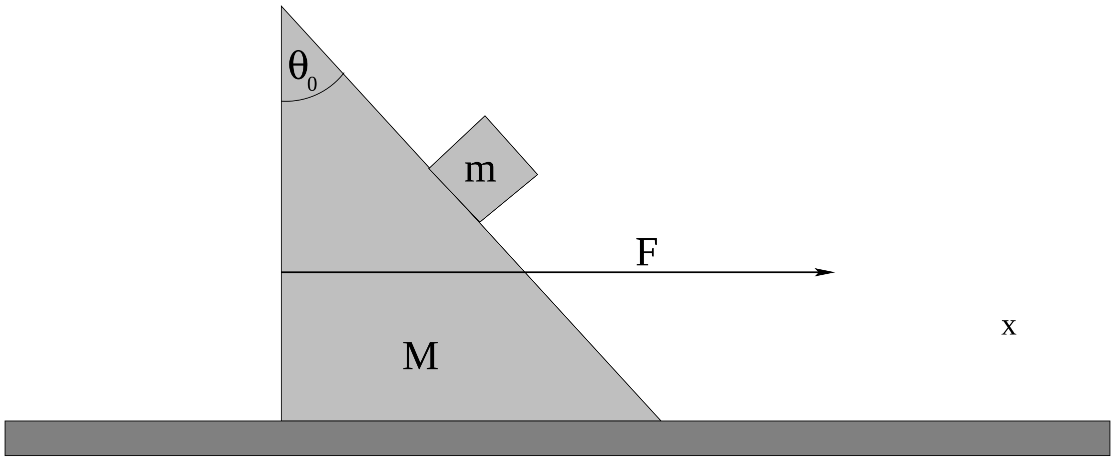
a) Draw the force/free body diagram for this problem.

b) Find the acceleration of the two masses.

c) Find the tension T in the string.

d) How fast are the two blocks moving when mass m_2 has fallen a height H (assuming that m_1 hasn't yet hit the pulley)?

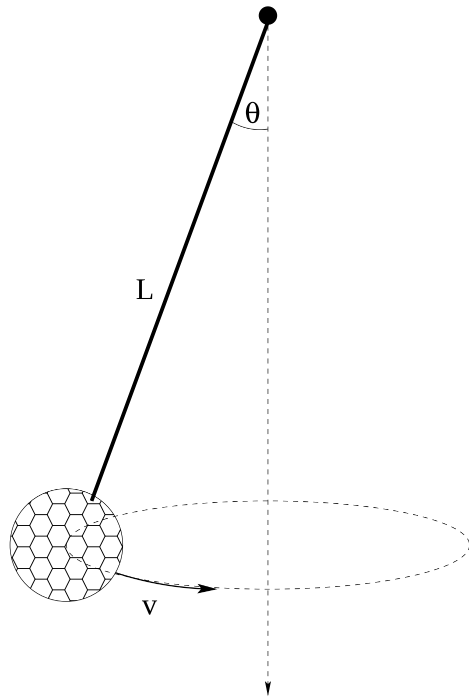
Problem 10.



A block m is sitting on a frictionless inclined block with mass M at an angle θ_0 as shown. With what force F should you push on the large block in order that the small block will remain motionless with respect to the large block and neither slide up nor slide down?

BTW, I made the angle θ_0 sit in the *upper* corner of the inclined plane just to make you *read the problem carefully* and actually *think* about sines and cosines of angles. Don't assume that every problem with an inclined plane will have the same (given) angles in the same places!

Problem 11.



A tether ball of mass m is suspended by a rope of length L from the top of a pole. A youngster gives it a whack so that it moves with some speed v in a circle of radius $r = L\sin(\theta) < L$ around the pole.

a) Find an expression for the tension T in the rope as a function of m, g, and θ.

b) Find an expression for the speed v of the ball as a function of θ.

Conceptual questions:

c) Why don't L and v appear in your (correct) answer for the tension T?

d) What is the limiting form of v for small angles (i.e. $\theta \ll 1$)?

Problem 12.

For this problem, you may find the picture above of a train's wheels useful. Note that they have a *rim* that *fits into* the rails on the inside. This thin rim of metal that rides on the inside of each rail is essential to the train being able to go around a curve and stay on a track! The purpose of this problem is for you to figure out *why* the wheels have this inset rim.

Suppose a train engine is chugging its way around a circular curve of radius R at a constant speed v, supported by its wheels sitting on the rail. We will assign the share of its mass carried by an outer wheel the symbol M_o and similarly assign M_i to the mass share of an inner wheel.

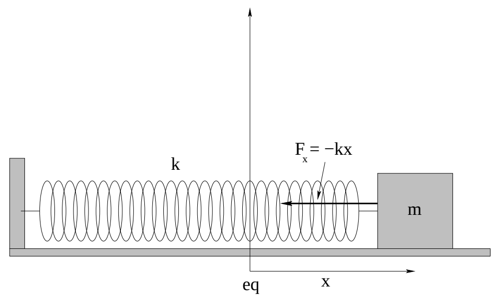
a) Draw a free body/force diagram for a pair of the train engine wheels (I'd suggest in cross-section) showing *all* of the forces acting on the inner and outer wheels *due to the rails* only as they bear their share of the weight of the engine. You diagram shoud clearly illustrate how a rail can exert *both* components of the force needed to hold a train up *and* curve its trajectory around in a circle.

b) Evaluate the *total vector force* acting on the pair of wheels together as a function of its speed in a plane perpendicular to its velocity $\vec{v}$. (That is, write N2 for y and x, using what you know about circular motion, gravity, and forces between rails and wheel.)

c) What is the mechanical origin of the force responsible for making the train go in a curve without coming off of the track (and for that matter, keeping it on the track in the first place, even when it is going "straight")?

d) What would happen if there were no rim on the train's wheels?

Problem 13.



A mass m on a frictionless table is connected to a spring with spring constant k (so that the force on it is $F_x = -kx$ where x is the distance of the mass from its equilibrium postion). It is then pulled so that the spring is stretched by a distance X_0 from its equilibrium position and at $t = 0$ is released.

a) Write Newton's Second Law and solve for the acceleration at a time it is at the arbitrary position x. Solve for the acceleration and write the result as a *second order, homogeneous differential equation* of motion for this system. Look ahead at chapter 9 if you need a hint to see how to do this.

b) Based on your life experience with masses on springs, how do you expect the mass to move in time? Again, look ahead at chapter 9 if necessary to see how to describe the motion mathematically.

c) Since $x(t)$ is not constant and a is proportional to $x(t)$, $\boldsymbol{a}$ **is a function of time**! Do you expect the solution to resemble the kinds of solutions you derived in constant acceleration problems above?

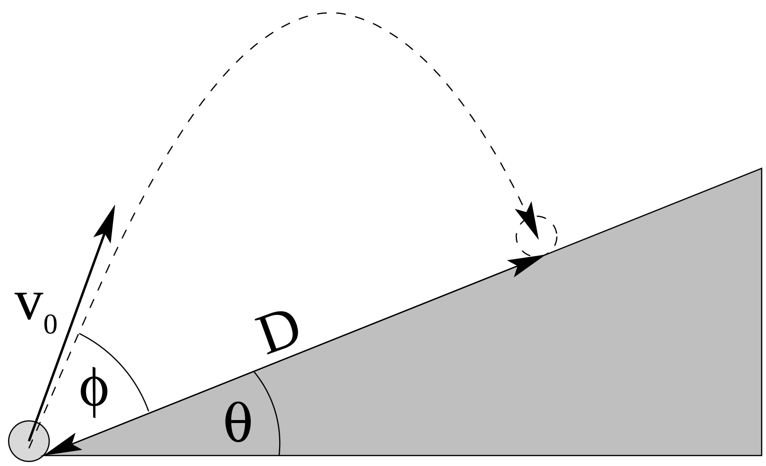
The point of this problem is for you to see right now, at the very beginning of the course while still working on the calculus of constant force and constant acceleration, that *not everything moves under the influence of a constant total force!* In fact, if anything, constant forces are rare, the exception rather than the rule, although doing the homework for the first few chapters or taking a high school physics class might convince you otherwise.

Note well: If the acceleration of a particle *varies in time*, we *cannot* use e.g. the constant acceleration solution:

$$x(t) = \frac{1}{2}at^2 + v_0t + x_0$$

Yet doing so is a **very common mistake** made by intro physics students who fail to realize this, sometimes as late as the final exam. Try to make sure that ***you are not one of them!***

Advanced Problem 14.



A ball of mass m is launched with an initial speed v_0 at an angle ϕ ***measured from the surface of an incline*** as shown above, which *itself* makes an angle θ with the ground. Ignoring drag – once fired the ball experiences only the force of gravity:

a) Find the distance D, measured ***along the incline***, from the launch point to where the ball strikes the incline.

b) What angle ϕ gives the maximum distance D?

Hints/Discussion: Note that there are two somewhat different approaches to this problem – both can be made to work. Both require the use of stuff like e.g. trig identities, the quadratic formula, and some clever way of identifying "the ball intercepts the slope at some specific time". Good luck!

Week 2: Newton's Laws: Continued

1.8: Summary

We now continue our discussion of dynamics and Newton's Laws, adding a few more very important force rules to our repertoire. So far our idealizations have carefully excluded forces that bring things to *rest* as they move, forces that always seem to act to *slow things down* unless we constantly push on them. The *dissipative* forces are, of course, ubiquitous and we cannot afford to ignore them for long. We'd also like to return to the issue of inertial reference frames and briefly discuss the topic of *pseudoforces* introduced in the "weight in an elevator" example above. Naturally, we will also see many examples of the use of these ideas, and will have to do even *more* problems for homework to make them intelligible.

The ideas we will cover include:

- **Static Friction** [66] is the force exerted by one surface on another that acts *parallel* to the surfaces to *prevent the two surfaces from sliding*.
  a) Static friction is as large as it needs to be to prevent any sliding motion, *up to* a maximum value, at which point the surfaces begin to slide.
  b) The maximum force static friction can exert is proportional to *both* the pressure between the surfaces *and* the area in contact. This makes it proportional to the product of the pressure and the area, which equals the normal force. We write its magnitude as:
  $$0 \leq f_s < \mu_s N \tag{2.1}$$
  where μ_s is the **coefficient of static friction**, a dimensionless constant characteristic of the two surfaces in contact, and N is the normal force.
  Note that we will hold the convention that f_s is ***never equal to*** $\boldsymbol{\mu_s N}$ – it is strictly less than this quantity in magnitude. This makes $\mu_s N$ the magnitude of the total shear force parallel to the surface neeed to *overcome* static friction and cause the two surfaces to slide. It also means that f_s in most problems is an *unknown* that must be solved for in the course of solving the problem, not a quantity that is known as soon as you known N.
  c) The direction of static friction is parallel to the surfaces in contact and *opposes* the component in that plane of the total force acting on the object (not including static

[66] Wikipedia: http://www.wikipedia.org/wiki/Friction. This article describes some aspects of friction in more detail than my brief introduction below. The standard model of friction I present is at best an approximate, idealized one. Wikipedia: http://www.wikipedia.org/wiki/Tribology describes the science of friction and lubrication in more detail.

friction) so that the total force parallel to the plane is *zero*. Note that in this course it will not matter which direction the applied force points in the plane of contact – static friction will act symmetrically to the right or left, backwards or forwards as needed to hold an object stationary.

- **Kinetic Friction** is the force exerted by one surface on another that is *sliding* across it. It, also, acts *parallel* to the surfaces and *opposes the direction of relative motion of the two surfaces*. That is:

  a) The force of kinetic friction is proportional to *both* the pressure between the surfaces *and* the area in contact. This makes it proportional to the product of the pressure and the area, which equals the normal force. Thus again

  $$f_k = \mu_k N \tag{2.2}$$

  where μ_k is the **coefficient of kinetic friction**, a dimensionless constant characteristic of the two surfaces in contact, and N is the normal force.

  Note well that kinetic friction *equals* $\mu_k N$ in magnitude, where static friction is whatever it needs to be to hold the surfaces static *up to* a maximum of $\mu_s N$. This is often a point of confusion for students when they first start to solve problems.

  b) The direction of kinetic friction is parallel to the surfaces in contact and *opposes* the ***relative direction of the sliding surfaces***. That is, if the bottom surface has a velocity (in any frame) of $\vec{v}_b$ and the top frame has a velocity of $\vec{v}_t \neq \vec{v}_b$, the direction of kinetic friction on the top object is the same as the direction of the vector $-(\vec{v}_t - \vec{v}_b) = \vec{v}_b - \vec{v}_t$. The bottom surface "drags" the top one in the (relative) direction it slides, as it were (and vice versa).

  Note well that *often* the circumstances where you will solve problems involving kinetic friction will involve a stationary lower surface, e.g. the ground, a fixed inclined plane, a roadway – all cases where kinetic friction simply opposes the direction of motion of the upper object – but you will be given enough problems where the lower surface is moving and "dragging" the upper one that you should be able to learn to manage them as well.

- **Drag Force**[67] is the "frictional" force exerted by a fluid (liquid or gas) on an object that moves through it. Like kinetic friction, it always opposes the direction of *relative* motion of the object and the medium: "drag force" equally well describes the force exerted on a car by the still air it moves through and the force exerted on a stationary car in a wind tunnel.

  Drag is an extremely complicated force. It depends on a vast array of things including but not limited to:

  - The size of the object.
  - The shape of the object.
  - The relative velocity of the object through the fluid.

[67]Wikipedia: http://www.wikipedia.org/wiki/Drag (physics). This article explains a *lot* of the things we skim over below, at least in the various links you can follow if you are particularly interested.

- The state of the fluid (e.g. its internal turbulence).
- The density of the fluid.
- The *viscosity* of the fluid (we will learn what this is later).
- The properties and chemistry of the surface of the object (smooth versus rough, strong or weak chemical interaction with the fluid at the molecular level).
- The *orientation* of the object as it moves through the fluid, which may be fixed in time or varying in time (as e.g. an object tumbles).

The long and the short of this is that actually computing drag forces on actual objects moving through actual fluids is a serious job of work for fluid engineers and physicists. To obtain mastery in this, one must first study for years, although then one can make a lot of money (and have a lot of fun, I think) working on cars, jets, turbine blades, boats, and many other things that involve the utilization or minimization of drag forces in important parts of our society.

To simplify drag forces to where we learn to understand *in general* how they work, we will use following idealizations:

a) We will only consider smooth, uniform, nonreactive surfaces of convex, bluff objects (like spheres) or streamlined objects (like rockets or arrows) moving through uniform, stationary fluids where we can ignore or treat separately e.g. bouyant forces.

b) We will wrap up all of our ignorance of the shape and cross-sectional area of the object, the density and viscosity of the fluid, and so on in a single number, b. This dimensioned number will only be actually computable for certain particularly "nice" shapes and so on (see the Wikipedia article on drag linked above) but allows us to treat drag relatively simply. We will treat drag in two limits:

c) Low velocity, non-turbulent (streamlined, laminar) motion leads to *Stokes' drag*, described by:
$$\vec{F}_d = -b\vec{v} \tag{2.3}$$
This is the simplest sort of drag – a drag force directly proportional to the velocity of (relative) motion of the object through the fluid and oppositely directed.

d) High velocity, turbulent (high Reynolds number) drag that is described by a *quadratic* dependence on the relative velocity:
$$\vec{F}_d = -b|v|\vec{v} \tag{2.4}$$
It is still directed opposite to the relative velocity of the object and the fluid but now is proportional to that velocity squared.

e) In between, drag is a bit of a mess – changing over from one from to the other. We will ignore this transitional region where turbulence is *appearing* and so on, except to note that it is there and you should be aware of it.

- **Pseudoforces in an accelerating frame** are gravity-like "imaginary" forces we must add to the real forces of nature to get an accurate Newtonian description of motion in a non-inertial reference frame. In all cases it is possible to solve Newton's Laws without recourse to pseudoforces (and this is the general approach we promote in this textbook) but it is useful in a few cases to see how to proceed to solve or formulate a problem using pseudoforces such as "centrifugal force" or "coriolis force" (both arising in a rotating frame) or pseudogravity in a linearly accelerating frame. In all cases if one tries to solve force equations in an accelerating frame, one must modify the actual force being exerted on a mass m in an inertial frame by:

$$\vec{F}_{\rm accelerating} = \vec{F}_{\rm inertial} - m\vec{a}_{\rm frame} = \vec{F}_{\rm inertial} + \vec{F}_p \tag{2.5}$$

where $\vec{F}_p = -m\vec{a}_{\rm frame}$ is the pseudoforce.

This sort of force is easily exemplified – indeed, we've already seen such an example in our treatment of apparent weight in an elevator in the first week/chapter.

2.1: Friction

So far, our picture of natural forces as being the *cause* of the acceleration of mass seems fairly successful. In time it will become second nature to you; you will intuitively connect forces to all changing velocities. However, our description thus far is fairly simplistic – we have *massless* strings, *frictionless* tables, *drag-free* air. That is, we are neglecting certain well-known and important facts or forces that appear in real-world problems in order to concentrate on "ideal" problems that illustrate the methods simply.

It is time to restore some of the complexity to the problems we solve. The first thing we will add is friction.

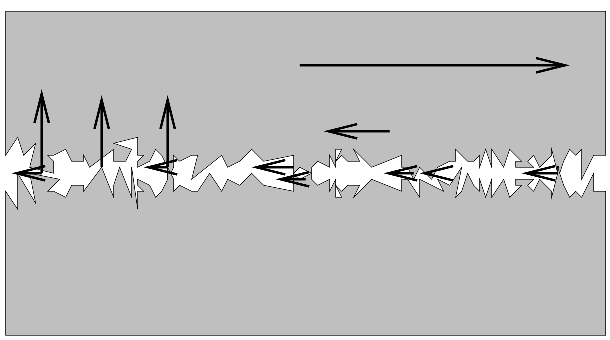


Figure 2.1: A cartoon picture representing two "smooth" surfaces in contact when they are highly magnified. Note the two things that contribute to friction – area in actual contact, which regulates the degree of chemical bonding between the surfaces, and a certain amount of "keyholing" where features in one surface fit into and are physically locked by features in the other.

Experimentally

a) $f_s \le \mu_s |N|$. The force exerted by *static friction* is less than or equal to the *coefficient of static friction* mu_s times the magnitude of the normal force exerted on the entire (homogeneous) surface of contact. We will sometimes refer to this maximum possible value of static friction as $f_s^{\max} = \mu_s |N|$. It opposes the component of any (otherwise net) applied force in the plane of the surface to make the total force component parallel to the surface zero as long as it is able to do so (up to this maximum).

b) $f_k = \mu_k |N|$. The force exerted by *kinetic friction* (produced by two surfaces rubbing against or sliding across each other in motion) is equal to the *coefficient of kinetic friction* times the magnitude of the normal force exerted on the entire (homogeneous) surface of contact. It opposes the direction of the *relative motion* of the two surfaces.

c) $\mu_k < \mu_s$

d) μ_k is really a function of the speed v (see discussion on drag forces), but for "slow" speeds $\mu_k \sim$ constant and we will idealize it as a constant throughout this book.

e) μ_s and μ_k depend on the materials in "smooth" contact, but are *independent* of contact area.

We can understand this last observation by noting that the frictional force should depend on the *pressure* (the normal force/area $\equiv$ N/m^2) *and* the area in contact. But then

$$f_k = \mu_k P * A = \mu_k \frac{N}{A} * A = \mu_k N \tag{2.6}$$

and we see that the frictional force will depend only on the total force, not the area or pressure separately.

The idealized force rules themselves, we see, are pretty simple: $f_s \leq \mu_s N$ and $f_k = \mu_k N$. Let's see how to apply them in the context of actual problems.

Example 2.1.1: Inclined Plane of Length L with Friction

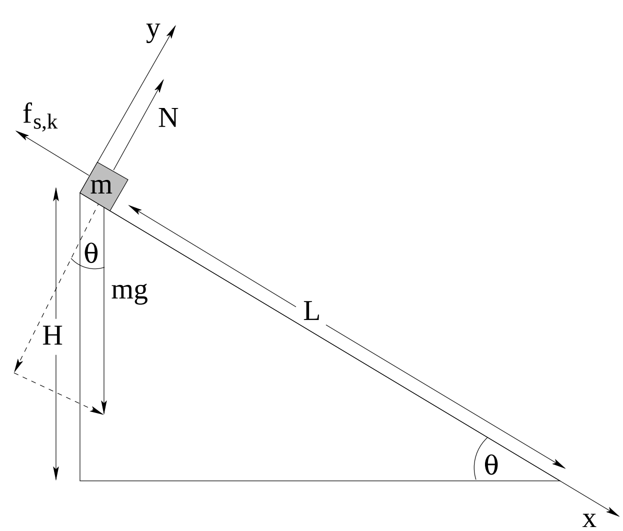


Figure 2.2: Block on inclined plane with both static and dynamic friction. Note that we still use the coordinate system selected in the version of the problem without friction, with the x-axis aligned with the inclined plane.

In figure 2.2 the problem of a block of mass m released from rest at time $t = 0$ on a plane of length L inclined at an angle θ relative to horizontal is once again given, this time more realistically, including the effects of *friction*. The inclusion of friction enables new questions to be asked that require the use of your knowledge of *both* the properties *and* the formulas that make up the friction force rules to answer, such as:

a) At what angle θ_c does the block *barely* overcome the force of static friction and slide down the incline.?

b) Started at rest from an angle $\theta > \theta_c$ (so it definitely slides), how fast will the block be going when it reaches the bottom?

To answer the first question, we note that static friction exerts *as much force as necessary* to keep the block at rest up to the maximum it can exert, $f_s^{\max} = \mu_s N$. We therefore decompose the known force rules into x and y components, sum them componentwise, write Newton's Second Law for both vector components and finally use our prior knowledge that the system remains in static force equilibrium to set $a_x = a_y = 0$. We get:

$$\sum F_x = mg\sin(\theta) - f_s = 0 \tag{2.7}$$

(for $\theta \leq \theta_c$ and $v(0) = 0$) and

$$\sum F_y = N - mg\cos(\theta) = 0 \tag{2.8}$$

So far, f_s is precisely what it needs to be to prevent motion:

$$f_s = mg\sin(\theta) \tag{2.9}$$

while

$$N = mg\cos(\theta) \tag{2.10}$$

is true at *any* angle, moving or not moving, from the F_y equation[68].

You can see that as one gradually and gently increases the angle θ, the force that must be exerted by static friction to keep the block in static force equilibrium increases as well. At the same time, the normal force exerted by the plane *decreases* (and hence the maximum force static friction *can* exert decreases as well. The critical angle is the angle where these two meet; where f_s is as large as it can be such that the block *barely* doesn't slide (or barely starts to slide, as you wish – at the boundary the slightest fluctuation in the total force suffices to trigger sliding). To find it, we can substitute $f_s^{\max} = \mu_s N_c$ where $N_c = mg\cos(\theta_c)$ into both equations, so that the first equation becomes:

$$\sum F_x = mg\sin(\theta_c) - \mu_s mg\cos(\theta_c) = 0 \tag{2.11}$$

at θ_c. Solving for θ_c, we get:

$$\theta_c = \tan^{-1}(\mu_s) \tag{2.12}$$

[68]Here again is an appeal to experience and intuition – we know that masses placed on inclines under the influence of gravity generally do not "jump up" off of the incline or "sink into" the (solid) incline, so their acceleration in the perpendicular direction is, from sheer common sense, zero. *Proving* this in terms of *microscopic interactions* would be absurdly difficult (although in principle possible) but as long as we keep our wits about ourselves we don't have to!

Once it is moving (either at an angle $\theta > \theta_c$ or at a smaller angle than this but with the initial condition $v_x(0) > 0$, giving it an initial "push" down the incline) then the block will (probably) *accelerate* and Newton's Second Law becomes:

$$\sum F_x = mg\sin(\theta) - \mu_k mg\cos(\theta) = ma_x \tag{2.13}$$

which we can solve for the constant acceleration of the block down the incline:

$$a_x = g\sin(\theta) - \mu_k g\cos(\theta) = g(\sin(\theta) - \mu_k\cos(\theta)) \tag{2.14}$$

Given a_x, it is now straightforward to answer the second question above (or any of a number of others) using the methods exemplified in the first week/chapter. For example, we can integrate twice and find $v_x(t)$ and $x(t)$, use the latter to find the time it takes to reach the bottom, and substitute that time into the former to find the speed at the bottom of the incline. Try this on your own, and get help if it isn't (by now) pretty easy.

Other things you might think about: Suppose that you started the block at the top of an incline at an angle *less* than θ_c but at an initial speed $v_x(0) = v_0$. In that case, it might well be the case that $f_k > mg\sin(\theta)$ and the block would slide down the incline *slowing down*. An interesting question might then be: Given the angle, μ_k, L and v_0, does the block come to rest before it reaches the bottom of the incline? Does the answer depend on m or g? Think about how you might formulate and answer this question in terms of the givens.

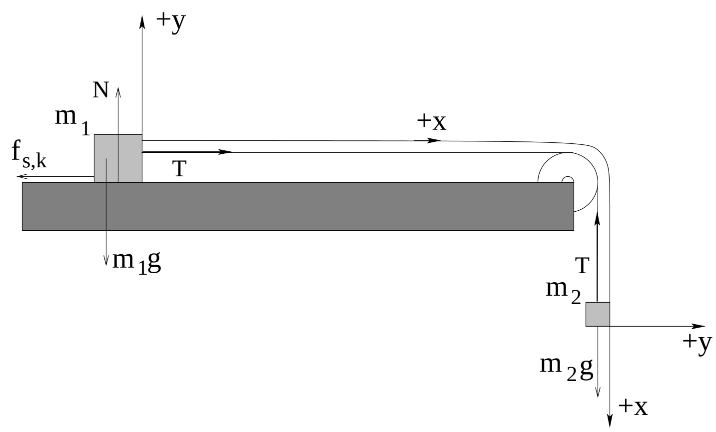


Figure 2.3: Atwood's machine, sort of, with one block resting on a table with friction and the other dangling over the side being pulled down by gravity near the Earth's surface. Note that we should use an "around the corner" coordinate system as shown, since $a_1 = a_2 = a$ if the string is unstretchable.

Suppose a block of mass m_1 sits on a table. The coefficients of static and kinetic friction between the block and the table are $\mu_s > \mu_k$ and μ_k respectively. This block is attached by an "ideal" massless unstretchable string running over an "ideal" massless frictionless pulley to a block of mass m_2 hanging off of the table as shown in figure 2.3. The blocks are released from rest at time $t = 0$.

Possible questions include:

a) What is the largest that m_2 can be before the system starts to move, in terms of the givens and knowns (m_1, g, μ_k, μ_s...)?

b) Find this largest m_2 if $m_1 = 10$ kg and $\mu_s = 0.4$.

c) Describe the subsequent motion (find a, $v(t)$, the displacement of either block $x(t)$ from its starting position). What is the tension T in the string while they are stationary?

d) Suppose that $m_2 = 5$ kg and $\mu_k = 0.3$. How fast are the masses moving after m_2 has fallen one meter? What is the tension T in the string while they are moving?

Note that this is the *first example* you have been given with actual numbers. They are there to tempt you to use your calculators to solve the problem. *Do not do this!* Solve *both* of these problems *algebraically* and only at the very end, with the full algebraic answers obtained and dimensionally checked, consider substituting in the numbers where they are given to get a numerical answer. In most of the rare cases you are given a problem with actual numbers in this book, they will be simple enough that you *shouldn't need* a calculator to answer them! Note well that the right number answer is worth *very little* in this course – I assume that all of you can, if your lives (or the lives of others for those of you who plan to go on to be physicians or aerospace engineers) depend on it, can punch numbers into a calculator correctly. This course is intended to teach you how to correctly obtain the algebraic expression that you need to numerically evaluate, not "drill" you in calculator skills[69].

We start by noting that, like Atwood's Machine and one of the homework problems from the first week, this system is effectively "one dimensional", where the string and pulley serve to "bend" the contact force between the blocks around the corner without loss of magnitude. I crudely draw such a coordinate frame into the figure, but bear in mind that it is really lined up with the string. The important thing is that the *displacement* of both blocks from their initial position is the same, and neither block moves perpendicular to "x" in their (local) "y" direction.

At this point the ritual should be quite familiar. For the first (static force equilibrium) problem we write Newton's Second Law with $a_x = a_y = 0$ for both masses and use *static* friction to describe the frictional force on m_1:

$$\begin{aligned}
\sum F_{x1} &= T - f_s = 0 \\
\sum F_{y1} &= N - m_1 g = 0 \\
\sum F_{x2} &= m_2 g - T = 0 \\
\sum F_{y2} &= 0
\end{aligned} \tag{2.15}$$

From the second equation, $N = m_1 g$. At the point where m_2 is the largest it can be (given m_1 and so on) $f_s = f_s^{\max} = \mu_s N = \mu_s m_1 g$. If we substitute this in and add the two x equations, the T cancels and we get:

$$m_2^{\max} g - \mu_s m_1 g = 0 \tag{2.16}$$

Thus

$$m_2^{\max} = \mu_s m_1 \tag{2.17}$$

which (if you think about it) makes both dimensional and physical sense. In terms of the given numbers, $m_2 > mu_s m_1 = 4$ kg is enough so that the weight of the second mass will make the

[69] Indeed, numbers are used as rarely as they are to *break* you of the bad habit of thinking that a calculator, or computer, is capable of doing your intuitive and formal algebraic reasoning for you, and are only included from time to time to give you a "feel" for what *reasonable* numbers are for describing everyday things.

whole system move. Note that the tension $T = m_2 g = 40$ Newtons, from F_{x2} (now that we know m_2).

Similarly, in the second pair of questions m_2 is larger than this minimum, so m_1 will *slide* to the right as m_2 falls. We will have to solve Newton's Second Law for both masses in order to obtain the *non*-zero acceleration to the right and down, respectively:

$$\begin{aligned}
\sum F_{x1} &= T - f_k = m_1 a \\
\sum F_{y1} &= N - m_1 g = 0 \\
\sum F_{x2} &= m_2 g - T = m_2 a \\
\sum F_{y2} &= 0
\end{aligned} \tag{2.18}$$

If we substitute the *fixed* value for $f_k = \mu_k N = \mu_k m_1 g$ and then add the two x equations once again (using the fact that both masses have the same acceleration because the string is unstretchable as noted in our original construction of round-the-corner coordinates), the tension T cancels and we get:

$$m_2 g - \mu_k m_1 g = (m_1 + m_2) a \tag{2.19}$$

or

$$a = \frac{m_2 - \mu_k m_1}{m_1 + m_2} g \tag{2.20}$$

is the constant acceleration.

This makes *sense!* The string forms an "internal force" not unlike the molecular forces that glue the many tiny components of *each* block together. As long as the two move together, these internal forces do not contribute to the collective motion of the system any more than you can pick yourself up by your own shoestrings! The net force "along x" is just the weight of m_2 pulling one way, and the force of kinetic friction pulling the other. The sum of these two forces equals the total mass times the acceleration!

Solving for $v(t)$ and $x(t)$ (for either block) should now be easy and familiar. So should finding the time it takes for the blocks to move one meter, and substituting this time into $v(t)$ to find out how fast they are moving at this time. Finally, one can substitute a into *either* of the two equations of motion involving T and solve for T. In general you should find that T is *less than* the weight of the second mass, so that the net force on this mass is not zero and accelerates it downward. The tension T can never be negative (as drawn) because strings can never push an object, only pull.

Basically, we are done. We know (or can easily compute) anything that can be known about this system.

Example 2.1.2: Find The Minimum No-Skid Braking Distance for a Car

One of the most important everyday applications of our knowledge of static versus kinetic friction is in *anti-lock brake systems* (ABS)[70] ABS brakes are implemented in every car sold in

[70] Wikipedia: http://www.wikipedia.org/wiki/Anti-lock Braking System.

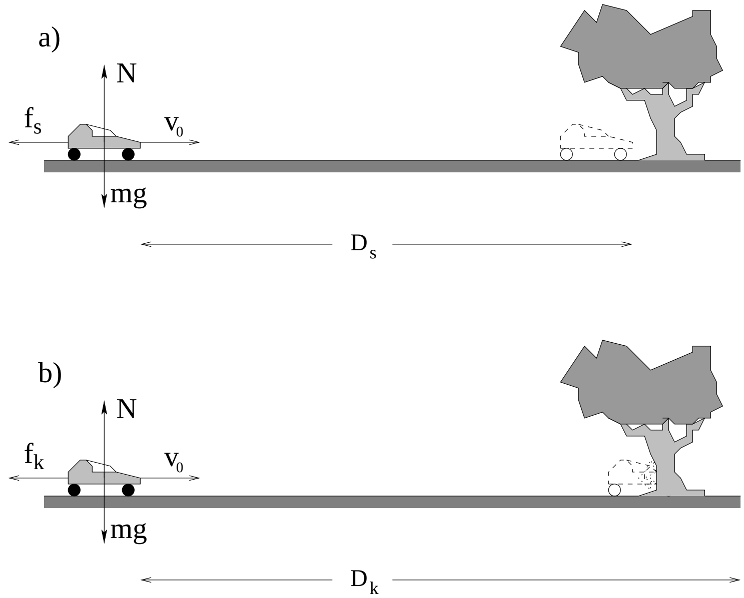


Figure 2.4: Stopping a car with and without locking the brakes and skidding. The coordinate system (not drawn) is x parallel to the ground, y perpendicular to the ground, and the origin in both cases is at the point where the car begins braking. In panel a), the anti-lock brakes do not lock and the car is stopped with the maximum force of static friction. In panel b) the brakes lock and the car skids to a stop, slowed by kinetic/sliding friction.

the European Union (since 2007) and are standard equipment in *almost* every car sold in the United States, where for reasons known only to congress it has yet to be formally mandated. This is in spite of the fact that road tests show that on average, stopping distances for ABS-equipped cars are some 18 to 35% shorter than non-ABS equipped cars, for all but the most skilled drivers (who still find it difficult to actually beat ABS stopping distances but who can equal them).

One small part of the reason may be that ABS braking "feels strange" as the car pumps the brakes *for* you 10-16 times per second, making it "pulse" as it stops. This causes drivers unprepared for the feeling to back off of the brake pedal and not take full advantage of the ABS feature, but of course the simpler and better solution is for drivers to educate themselves on the feel of anti-lock brakes in action under safe and controlled conditions and then *trust them*.

This problem is designed to help you *understand* why ABS-equipped cars are "better" (safer) than non-ABS-equipped cars, and why you should rely on them to help you stop a car in the minimum possible distance. We achieve this by answer the following questions:

Find the minimum braking distance of a car travelling at speed v_0 30 m/sec running on tires with $\mu_s = 0.5$ and $\mu_k = 0.3$:

a) equipped with ABS such that the tires do not skid, but rather roll (so that they exert the maximum static friction only);

b) the same car, but without ABS and with the wheels locked in a skid (kinetic friction only)

c) Evaluate these distances for $v_0 = 30$ meters/second (~ 67 mph), and both for $\mu_s = 0.8$, $\mu_k = 0.7$ (reasonable values, actually, for good tires on dry pavement) and for $\mu_s = 0.7$,

$\mu_k = 0.3$ (not unreasonable values for *wet* pavement). The latter are, however, highly variable, depending on the kind and conditions of the treads on your car (which provide channels for water to be displaced as a thin film of water beneath the treads lubricates the point of contact between the tire and the road. With luck they will teach you *why* you should slow down and allow the distance between your vehicle and the next one to stretch out when driving in wet, snowy, or icy conditions.

To answer all of these questions, it suffices to evalute the acceleration of the car given either $f_s^{\max} = \mu_s N = \mu_s mg$ (for a car being stopped by peak static friction via ABS) and $f_k = \mu_k N = \mu_k mg$. In both cases we use Newton's Law in the x-direction to find a_x:

$$\sum_x F_x = -\mu_{(s,k)} N = ma_x \quad (2.21)$$

$$\sum_y F_y = N - mg = ma_y = 0 \quad (2.22)$$

(where μ_s is for static friction and μ_k is for kinetic friction), or:

$$ma_x = -\mu_{(s,k)} mg \quad (2.23)$$

so

$$a_x = -\mu_{(s,k)} g \quad (2.24)$$

which is a constant.

We can then easily determine how long a distance D is required to make the car come to rest. We do this by finding the *stopping time* t_s from:

$$v_x(t_s) = 0 = v_0 - \mu_{(s,k)} g t_s \quad (2.25)$$

or:

$$t_s = \frac{v_0}{\mu_{(s,k)} g} \quad (2.26)$$

and using it to evaluate:

$$D_{(s,k)} = x(t_s) = -\frac{1}{2}\mu_{(s,k)} g t_s^2 + v_0 t_s \quad (2.27)$$

I will leave the actual completion of the problem up to you, because doing these last few steps four times will provide you with a valuable lesson that we will exploit shortly to motivate learing about *energy*, which will permit us to answer questions like this *without* always having to find times as intermediate algebraic steps.

Note well! The answers you obtain for D (if correctly computed) are *reasonable!* That is, yes, it can easily take you order of 100 meters to stop your car with an initial speed of 30 meters per second, and this doesn't even allow for e.g. reaction time. Anything that shortens this distance makes it easier to survive an emergency situation, such as avoiding a deer that "appears" in the middle of the road in front of you at night.

Example 2.1.3: Car Rounding a Banked Curve with Friction

A car of mass m is rounding a circular curve of radius R banked at an angle θ relative to the horizontal. The car is travelling at speed v (say, into the page in figure 2.5 above). The coefficient of static friction between the car's tires and the road is μ_s. Find:

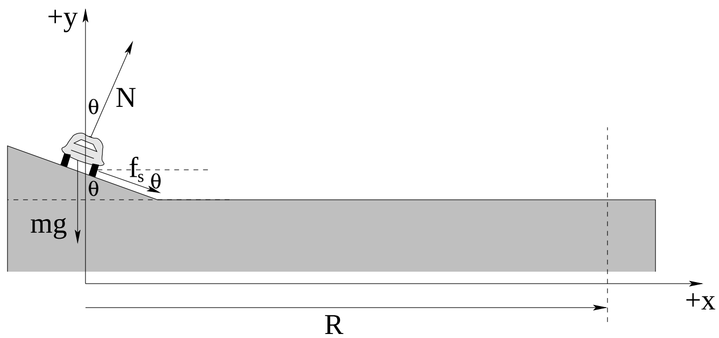


Figure 2.5: Friction and the normal force conspire to accelerate car towards the center of the circle in which it moves, together with the *best* coordinate system to use – one with one axis pointing in the direction of actual acceleration. Be sure to choose the right coordinates for this problem!

a) The normal force exerted by the road on the car.

b) The force of static friction exerted by the road on the tires.

c) The *range* of speeds for which the car can round the curve successfully (without sliding up or down the incline).

Note that we don't know f_s, but we are certain that it must be less than or equal to $\mu_s N$ in order for the car to successfully round the curve (the third question). To be able to formulate the range problem, though, we have to find the normal force (in terms of the other/given quantities and the force exerted by static friction (in terms of the other quantities), so we start with that.

As always, the only thing we really know is our dynamical principle – Newton's Second Law – plus our knowledge of the force rules involved plus *our experience and intuition*, which turn out to be *crucial* in setting up this problem.

For example, what direction should f_s point? Imagine that the inclined roadway is coated with frictionless ice and the car is sitting on it (almost) at rest (for a finite but tiny $v \to 0$). What will happen (if $\mu_s = \mu_k = 0$)? Well, obviously it will *slide down the hill* which doesn't qualify as 'rounding the curve' at a constant height on the incline. Now imagine that the car is travelling at an enormous v; what will happen? The car will skid off of the road to the outside, of course. We know (and fear!) that from our own experience rounding curves too fast.

We now have two *different* limiting behaviors – in the first case, to round the curve friction has to keep the car from sliding *down* at low speeds and hence must point *up* the incline; in the second case, to round the curve friction has to point *down* to keep the car from skidding up and off of the road.

We have little choice but to pick one of these two possibilities, solve the problem for that possibility, and then solve it *again* for the other (which should be as simple as changing the sign of f_s in the algebra. I therefore arbitrarily picked f_s pointing *down* (and parallel to, remember) the incline, which will eventually give us the *upper* limit on the speed v with which we can round the curve.

As always we use coordinates lined up with the eventual direction of $\vec{F}_{tot}$ and the actual acceleration of the car: $+x$ parallel to the *ground* (and the plane of the circle of movement with radius R).

We write Newton's second law:

$$\sum_x F_x = N\sin\theta + f_s\cos\theta = ma_x = \frac{mv^2}{R} \tag{2.28}$$

$$\sum_y F_y = N\cos\theta - mg - f_s\sin\theta = ma_y = 0 \tag{2.29}$$

(where so far f_s is not its maximum value, it is merely whatever it needs to be to make the car round the curve for a v presumed to be in range) and solve the y equation for N:

$$N = \frac{mg + f_s\sin\theta}{\cos\theta} \tag{2.30}$$

substitute into the x equation:

$$(mg + f_s\sin\theta)\tan\theta + f_s\cos\theta = \frac{mv^2}{R} \tag{2.31}$$

and finally solve for f_s:

$$f_s = \frac{\frac{mv^2}{R} - mg\tan\theta}{\sin\theta\tan\theta + \cos\theta} \tag{2.32}$$

From this we see that if

$$\frac{mv^2}{R} > mg\tan\theta \tag{2.33}$$

or

$$\frac{v^2}{Rg} > \tan\theta \tag{2.34}$$

then f_s is positive (down the incline), otherwise it is negative (up the incline). When $\frac{v^2}{Rg} = \tan\theta$, $f_s = 0$ and the car would round the curve even on ice (as you determined in a previous homework problem).

See if you can use your knowledge of the algebraic form for $f_s^{\max}$ to determine the range of v *given* μ_s that will permit the car to round the curve. It's a bit tricky! You may have to go back a couple of steps and find $N^{\max}$ (the N associated with $f_s^{\max}$) *and* $f_s^{\max}$ in terms of that N at the same time, because both N and f_s depend, in the end, on v...

2.2: Drag Forces

As we will discuss later in more detail in the week that we cover fluids, when an object is sitting at rest in a fluid at rest with a uniform temperature, pressure and density, the fluid around it presses on it, on average, equally on all sides[71].

Basically, the molecules of the fluid on one side of the object hit it, on average, with as much force per unit area area as molecules on the other side and the total cross-sectional area of the object seen from any given direction or the opposite of that direction is the same.

[71]We are ignoring variations with bulk fluid density and pressure in e.g. a gravitational field in this *idealized* statement; later we will see how the field gradient gives rise to ***buoyancy*** through ***Archimedes' Principle***. However, lateral forces perpendicular to the gravitational field and pressure gradient still cancel even then.

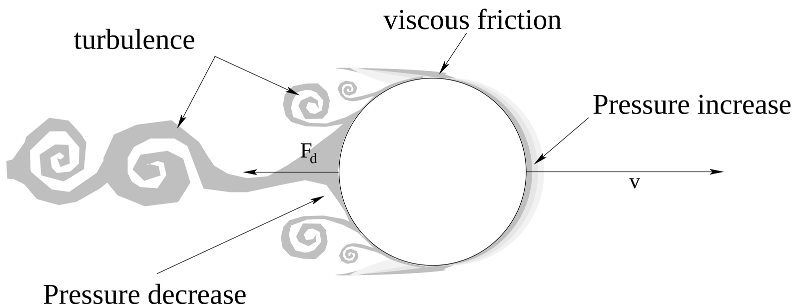


Figure 2.6: A "cartoon" illustrating the differential force on an object moving through a fluid. The drag force is associated with a differential pressure where the pressure on the side facing into the 'wind' of its passage is higher than the pressure of the trailing/lee side, plus a "dynamic frictional" force that comes from the fluid rubbing on the sides of the object as it passes. In very crude terms, the former is proportional to the cross-sectional area; the latter is proportional to the surface area exposed to the flow. However, the *details* of even this simple model, alas, are enormously complex.

By the time one works out all of the vector components and integrates the force component along any line over the whole surface area of the object, the force cancels. This "makes sense" – the whole system is in average static force equilibrium and we don't *expect* a tree to bend in the wind when there is no wind!

When the same object is *moving* with respect to the fluid (or the fluid is moving with respect to the object, i.e. – there is a wind in the case of air) then we empirically observe that a friction-like force is exerted on the object (and back on the fluid) called ***drag***[72] .

We can make up at least an heuristic description of this force that permits us to intuitively reason about it. As an object moves through a fluid, one expects that the molecules of the fluid will hit on the side *facing* the direction of motion harder, on the average, then molecules on the other side. Even though we will delay our formal treatment of fluid pressure until later, we should all be able to understand that these stronger collisions correspond (on average) to a greater *pressure* on the side of the object moving *against* the fluid or vice versa, and a lower pressure in the turbulent flow on the far side, where the object is moving away from the "chasing" and disarranged molecules of fluid. This pressure-linked drag force we might expect to be proportional to the *cross-sectional area* of the object perpendicular to its direction of relative motion through the fluid and is called ***form drag*** to indicate its strong dependence on the shape of the object.

However, the fluid that flows over the *sides* of the object also tends to "stick" to the surface of the object because of molecular interactions that occur during the instant of the molecular collision between the fluid and the surface. These collisions exert transverse "frictional" forces that tend to speed up the recoiling air molecules in the direction of motion of the object and slow the object down. The interactions can be strong enough to actually "freeze" a thin layer called the *boundary layer* of the fluid right up next to the object so that the frictional forces are transmitted through successive layers of fluid flowing and different speeds relative to the

[72]Wikipedia: http://www.wikipedia.org/wiki/Drag (physics). This is a nice summary and well worth at least glancing at to take note of the figure at the top illustrating the progression from laminar flow and skin friction to highly turbulent flow and pure form drag.

object. This sort of flow in layers is often called ***laminar*** (layered) flow and the frictional force exerted on the object transmitted through the rubbing of the layers on the sides of the object as it passes through the fluid is called ***skin friction*** or *laminar drag*.

Note well: When an object is enlongated and passes through a fluid parallel to its long axis with a comparatively small forward-facing cross section compared to its total area, we say that it is a ***streamlined object*** as the fluid tends to pass over it in laminar flow. A streamlined object will often have its total drag dominated by skin friction. A ***bluff object***, in contrast has a comparatively large cross-sectional surface facing forward and will usually have the total drag dominated by form drag. Note that a *single* object, such as an arrow or piece of paper, can often be streamlined moving through the fluid one way and bluff another way or be crumpled into a different shape with any mix in between. A sphere is considered to be a bluff body, dominated by form drag.

Unfortunately, this is only the *beginning* of an heuristic description of drag. Drag is a ***very complicated force***, especially when the object isn't smooth or convex but is rather rough and irregularly shaped, or when the fluid through which it moves is not in an "ideal" state to begin with, when the object itself *tumbles* as it moves through the fluid causing the drag force to constantly change form and magnitude. Flow over different parts of a single object can be laminar here, or ***turbulent*** there (with portions of the fluid left spinning in whirlpool-like eddies in the wake of the object after it passes).

The full Newtonian description of a moving fluid is given by the ***Navier-Stokes*** equation[73] which is too hard for us to even look at.

We will therefore need to idealize; learn a few nearly universal heuristic rules that we can use to conceptually understand fluid flow for at least simple, smooth, convex geometries.

It would be nice, perhaps, to be able to skip all of this but we can't, not even for future physicians as opposed to future engineers, physicists or mathematicians. As it happens, the body contains at least two major systems of fluid flow – the vasculature and the lymphatic system – as well as numerous minor ones (the renal system, various sexual systems, even much of the digestive system is at least partly a fluid transport problem). Drag forces play a critical role in understanding blood pressure, heart disease, and lots of other stuff. Sorry, my beloved students, you gotta learn it at least well enough to qualitatively and semi-quantitatively understand it.

Besides, this section is the key to *understanding* how to at least *in principle* fall out of an airplane without a parachute and survive. Drag forces *significantly modify* the idealized trajectory functions we derived in week 1, so much so that anyone relying on them to aim a cannon would almost certainly *consistently miss* any target they aimed at using the idealized no-drag trajectories.

Drag is an extremely complicated force. It depends on a vast array of things including but not limited to:

[73] Wikipedia: http://www.wikipedia.org/wiki/Navier-Stokes Equation. A partial differential way, way beyond the scope of this course. To give you an idea of how difficult the Navier-Stokes equation is to solve (in all but a few relatively simple geometries) simply demonstrating that solutions to it always *exist* and are *smooth* is one of the seven most important questions in mathematics and you could win a million dollar prize if you were to demonstrate it (or offer a proven counterexample).

- The size of the object.
- The shape of the object.
- The relative velocity of the object through the fluid.
- The state of the fluid (e.g. its velocity field including any internal turbulence).
- The density of the fluid.
- The *viscosity* of the fluid (we will learn what this is later).
- The properties and chemistry of the surface of the object (smooth versus rough, strong or weak chemical interaction with the fluid at the molecular level).
- The *orientation* of the object as it moves through the fluid, which may be fixed in time (streamlined versus bluff motion) or varying in time (as, for example, an irregularly shaped object tumbles).

To eliminate most of this complexity and end up with "force rules" that will often be quantitatively predictive we will use a number of idealizations. We will only consider smooth, uniform, nonreactive surfaces of convex bluff objects (like spheres) or streamlined objects (like rockets or arrows) moving through uniform, stationary fluids where we can ignore or treat separately the other non-drag (e.g. buoyant) forces acting on the object.

There are two dominant contributions to drag for objects of this sort.

The first, as noted above, is ***form drag*** – the difference in pressure times projective area between the front of an object and the rear of an object. It is strongly dependent on both the shape and orientation of an object and requires at least some turbulence in the trailing wake in order to occur.

The second is ***skin friction***, the friction-like force resulting from the fluid rubbing *across* the skin at right angles in laminar flow.

In this course, we will wrap up all of our *ignorance* of the shape and cross-sectional area of the object, the density and viscosity of the fluid, and so on into a *single number:* b. This (dimensioned) number will only be actually *computable* for certain particularly "nice" shapes, but it allows us to understand drag qualitatively and treat drag semi-quantitatively relatively simply in two important limits.

2.2.1: Stokes, or Laminar Drag

The first is when the object is moving through the fluid relatively slowly and/or is arrow-shaped or rocket-ship-shaped so that streamlined **laminar** drag (skin friction) is dominant (see figure 2.7). In this case there is relatively little form drag, and in particular, there is little or no ***turbulence*** – eddies of fluid spinning around an axis – in the wake of the object as the presence of turbulence (which we will discuss in more detail later when we consider fluid dynamics) breaks up laminar flow.

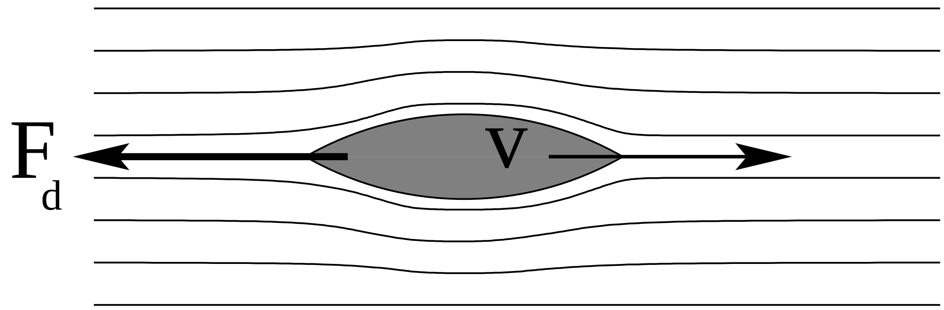


Figure 2.7: A streamlined object moving comparatively slowly through a fluid experiences *laminar* (Stokes) drag as the object displaces the fluid out of the way without initiating turbulent eddies.

This "low-velocity, streamlined" skin friction drag is technically named ***Stokes' drag*** (as Stokes was the first to derive it as a particular limit of the Navier-Stokes equation for a sphere moving through a fluid) or laminar drag and has the idealized force rule:

$$\vec{F}_d = -b\vec{v} \tag{2.35}$$

This is the *simplest sort of drag* – a drag force ***directly proportional to the velocity of relative motion of the object through the fluid and oppositely directed.***

Stokes derived the following relation for the dimensioned number b (the laminar drag coefficient) that appears in this equation for a *sphere* of radius R:

$$b = -6\pi\mu R \tag{2.36}$$

where μ is the *dynamical viscosity*. Different objects will have different laminar drag coefficients b, and in general it will be used as a simple given parameter in any problem involving Stokes drag.

Sadly – sadly because Stokes drag is remarkably mathematically tractable compared to e.g. turbulent drag below – spheres experience pure Stokes drag only when they are *very small* or moving *very slowly* through the fluid. To given you an idea of how slowly – a sphere moving at 1 meter per second through water would have to be on the order of one *micron* (a millionth of a meter) in size in order to experience predominantly laminar/Stokes drag. Equivalently, a sphere a meter in diameter would need to be moving at a micron per second. This is a force that is relevant for bacteria or red blood cells moving in water, but not too relevant to baseballs.

It becomes *more* relevant for streamlined objects – objects whose length *along* the direction of motion greatly exceeds the characteristic length of the cross-sectional area perpendicular to this direction. We will therefore still find it useful to solve a few problems involving Stokes drag as it will be highly relevant to our eventual studies of harmonic oscillation and is not irrelevant to the flow of blood in blood vessels.

2.2.2: Rayleigh, or Turbulent Drag

On the other hand, if one moves an object through a fluid *too fast* – where the actual speed depends in detail on the actual size and shape of the object, how bluff or streamlined it is –

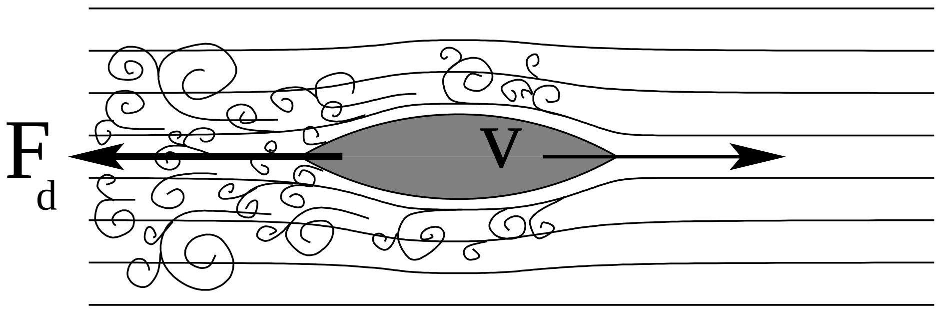


Figure 2.8: A streamlined object moving rapidly through a fluid experiences *turbulent* (Rayleigh) drag as the object imparts enough momentum to the surrounding fluid to initiate turbulent eddies and alter the *scaling* of the force with speed.

pressure builds up on the leading surface and ***turbulence***[74] appears in its trailing wake in the fluid (as illustrated in figure 2.8 above) when the *Reynolds number* R_e of the relative motion (which is a function of the relative velocity, the kinetic viscosity, and the characteristic length of the object)) exceeds a critical threshold. Again, we will learn more about this (and perhaps define the Reynolds number) later – for the moment it suffices to know that *most* macroscopic objects moving through water or air at reasonable velocities experience turbulent drag, not Stokes drag.

This high velocity, ***turbulent drag*** exerts a force described by a *quadratic* dependence on the relative velocity due to Lord Rayleigh:

$$\vec{\boldsymbol{F}}_d = -\frac{1}{2}\rho C_d A|v|\vec{\boldsymbol{v}} = -c|v|\vec{\boldsymbol{v}} \tag{2.37}$$

It is still ***directed opposite to the relative velocity of the object and the fluid*** but now is proportional to that velocity ***squared***. In this formula ρ is the density of the fluid through which the object moves (so denser fluids exert more drag as one would expect) and A is the *cross-sectional area* of the object *perpendicular to the direction of motion*, also known as the *orthographic projection* of the object on any plane perpendicular to the motion. For example, for a sphere of radius R, the orthographic projection is a circle of radius R and the area $A = \pi R^2$.

The number C_d is called the *drag coefficient* and is a dimensionless number that depends on relative speed, flow direction, object position, object size, fluid viscosity and fluid density. In other words, the expression above is only valid in certain domains of all of these properties where C_d is *slowly varying* and can be thought of as a "constant"! Hence we can say that for a sphere moving through still air at speeds where turbulent drag is dominant it is around $0.47 \approx 0.5$ and wrap all of the dependencies on the size of the sphere and density of the fluid into a single dimensioned number, c:

$$c \approx \frac{1}{4}\rho\pi R^2 \tag{2.38}$$

[74]Wikipedia: http://www.wikipedia.org/wiki/Turbulence. Turbulence – eddies spun out in the fluid as it moves off of the surface passing throughout it – is arguably the single most complex phenomenon physics attempts to describe, dwarfing even things like quantum field theory in its difficulty. We can "see" a great deal of structure in it, but that structure is fundamentally *chaotic* and hence subject to things like the ***butterfly effect***. In the end it is very difficult to compute except in certain limiting and idealized cases.

which one can then more or less directly compare to $b = 6\mu\pi R$ for the Stokes drag of the *same* sphere, moving much slower. b and c also allow us to study the semi-quantitative effect of the two kinds of drag in isolation, below, with only a single parameter characteristic of the object/fluid to vex us.

To get a feel for non-spherical objects, bluff convex objects like potatoes or cars or people have drag coefficients close to but a bit more or less than 0.5, while highly bluff objects might have a drag coefficient over 1.0 and truly streamlined objects might have a drag coefficient as low as 0.04.

As one can see, the functional complexity of the actual *non*-constant drag coefficient C_d even for such a simple object as a sphere has to manage the entire transition from laminar drag force for low velocities/Reynold's numbers to turbulent drag for high velocites/Reynold's numbers, so that at speeds in between the drag force is at best a function of a *non*-integer power of v in between 1 and 2 or some arcane mixture of form drag and skin friction. We will pretty much ignore this transition. It is just too damned difficult for us to mess with, although you should certainly be aware that it is *there*. We will also skip at this time the surprising order that can appear at stages in this transition as the turbulent rolls often can appear in a quite regular/periodic way before finally becoming chaotic in full turbulence. Nonlinear fluid dynamics, as noted above, is a *very difficult subject* and well beyond the scope of this initial introduction to some of its simpler results.

You can see that in our actual expression for the drag force above, as promised, we have simplified things *even more* and express all of this dependence – ρ, μ, size and shape and more – wrapped up in the turbulent dimensioned constant c, which one can think of as an overall turbulent "drag coefficient" that plays the same general role as the laminar "drag coefficient" b we similarly defined above. However, it is impossible for the heuristic descriptors b and c to be the *same* for Stokes' and turbulent drag – they don't even have the same *units* or dependencies on e.g. the size of the object or its shape – and for most objects most of the time the total drag is some sort of mixture of these limiting forces, with one or the other (probably) dominant.

Even so, these simplified forms turn out to be accurate for some important things we'd like to understand, and indeed are both quantitatively and qualitatively useful to us in a number of contexts where we can solve the equations of motion and graph the results, and use them to understand things like: terminal velocity of a falling object; the dynamics of water flowing in circular cross-section pipes; blood flowing in circular cross-section blood vessels; how the streamlined shape of a car can affect its gas mileage.

We will look at some of these things now, and postpone discussion of others to the chapter on the physics of fluids.

2.2.3: Terminal velocity

One immediate consequence of this is that objects dropped in a gravitational field in fluids such as air or water do not just keep speeding up *ad infinitum*. When they are dropped from rest, at first their speed is very low and drag forces may well be negligible[75]. The gravitational

[75]In air and other low viscosity, low density compressible gases they probably are; in water or other viscous, dense, incompressible liquids they may not be.

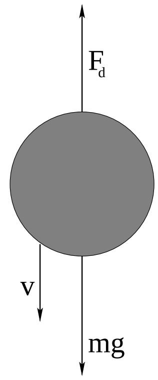


Figure 2.9: A simple object falling through a fluid experiences a drag force that increasingly cancels the force of gravity as the object accelerates until a *terminal velocity* v_t is asymptotically reached. For bluff objects such as spheres, the $F_d = -bv^2$ force rule is usually appropriate.

force accelerates them downward and their speed gradually increases.

As it increases, however, the drag force in all cases *increases as well*. For many objects the drag forces will quickly transition over to turbulent drag, with a drag force magnitude of cv^2. For others, the drag force may remain Stokes' drag, with a drag force magnitude of $b_l v$ (in both cases opposing the directioin of motion through the fluid). Eventually, the drag force will *balance* the gravitational force and the object will no longer accelerate. It will fall instead at a *constant speed*. This speed is called the *terminal velocity*.

It is extremely easy to compute the terminal velocity for a falling object, given the form of its drag force rule. It is the velocity where the net force on the object vanishes. If we choose a coordinate system with "down" being e.g. x positive (so gravity and the velocity are both positive pointing down) we can write either:

$$\begin{aligned} mg - bv_t &= ma_x = 0 \qquad \text{or} \\ v_t &= \frac{mg}{b} \end{aligned} \tag{2.39}$$

(for Stokes' drag) or

$$\begin{aligned} mg - cv_t^2 &= ma_x = 0 \qquad \text{and} \\ v_t &= \sqrt{\frac{mg}{c}} \end{aligned} \tag{2.40}$$

for turbulent drag.

We expect $v_x(t)$ to *asymptotically approach* v_t with time. Rather than draw a *generic* asymptotic curve (which is easy enough, just start with the slope of v being g and bend the curve over to smoothly approach v_t), we will go ahead and see how to solve the equations of motion for at least the two limiting (and common) cases of Stokes' and turbulent drag.

The entire complicated set of drag formulas above can be reduced to the following "rule of thumb" that applies to objects of water-like density that have sizes such that turbulent drag determines their terminal velocity – raindrops, hail, live animals (including humans) falling in air just above sea level near the surface of the Earth. In this case terminal velocity is *roughly* equal to

$$v_t = 90\sqrt{d} \tag{2.41}$$

where d is the characteristic size of the object in meters.

For a human body $d \approx 0.6$ so $v_t \approx 70$ meters per second or 156 miles per hour. However, if one falls in a bluff position, one can reduce this to anywhere from 40 to 55 meters per second, say 90 to 120 miles per hour.

Note that the characteristic size of a small animal such as a squirrel or a cat might be 0.05 (squirrel) to 0.1 (cat). Terminal velocity for a cat is around 28 meters per second, lower if the cat falls in a bluff position (say, 50 to 60 mph) and for a squirrel in a bluff position it might be as low as 10 to 20 mph. Smaller animals – especially ones with large bushy tails or skin webs like those observed in the *flying squirrel*[76] – have a much lower terminal velocity than (say) humans and hence have a much better chance of survival. One rather imagines that this provided a direct evolutionary path to actual flight for small animals that lived relatively high above the ground in arboreal niches.

Example 2.2.1: Falling From a Plane and Surviving

As noted above, the terminal velocity for humans in free fall near the Earth's surface is (give or take, depending on whether you are falling in a streamlined swan dive or falling in a bluff skydiver's belly flop position) anywhere from 40 to 70 meters per second (90-155 miles per hour). Amazingly, humans can survive[77] collisions at this speed.

The trick is to fall into something *soft and springy* that *gradually* slows you from high speed to zero without ever causing the deceleration force to exceed 100 times your weight, applied as uniformly as possible to parts of your body you can live without such as your legs (where your odds go up the smalller this multiplier is, of course). It is pretty simple to figure out what kinds of things might do.

Suppose you fall from a large height (long enough to reach terminal velocity) to hit a haystack of height H that exerts a nice, uniform force to slow you down all the way to the ground, smoothly compressing under you as you fall. In that case, your initial velocity at the top is v_t, down. In order to stop you before $y = 0$ (the ground) you have to have a net acceleration $-a$ such that:

$$v(t_g) = 0 = v_t - at_g \tag{2.42}$$

$$y(t_g) = 0 = H - v_t t_g - \frac{1}{2}at_g^2 \tag{2.43}$$

If we solve the first equation for t_g (something we have done many times now) and substitute it into the second and solve for the magnitude of a, we will get:

$$\begin{aligned} -v_t^2 &= -2aH \qquad \text{or} \\ a &= \frac{v_t^2}{2H} \end{aligned} \tag{2.44}$$

[76]Wikipedia: http://www.wikipedia.org/wiki/Flying Squirrel. A flying squirrel doesn't really fly – rather it skydives in a highly bluff position so that it can glide long transverse distances and land with a very low terminal velocity.

[77]Wikipedia: http://www.wikipedia.org/wiki/Free_fall#Surviving_falls. ...and *have* survived...

We know also that

$$F_{\rm haystack} - mg = ma \tag{2.45}$$

or

$$F_{\rm haystack} = ma + mg = m(a+g) = mg' = m\left(\frac{v_t^2}{2H} + g\right) \tag{2.46}$$

Let's suppose the haystack was $H = 1.25$ meter high and, because you cleverly landed on it in a "bluff" position to keep v_t as small as possible, you start at the top moving at only $v_t = 50$ meters per second. Then $g' = a + g$ is approximately 1009.8 meters/second^2, 103 'gees', and the force the haystack must exert on you is 103 times your normal weight. You actually have a small chance of surviving this stopping force, but it isn't a very large one.

To have a better chance of surviving, one needs to keep the g-force under 100, ideally *well* under 100, although a very few people are known to have survived 100 g accelerations in e.g. race car crashes. Since the "haystack" portion of the acceleration needed is inversely proportional to H we can see that a 2.5 meter haystack would lead to 51 gees, a 5 meter haystack would lead to 26 gees, and a 10 meter haystack would lead to a mere 13.5 gees, nothing worse than some serious bruising. If you want to get up and walk to your press conference, you need a haystack or palette at the mattress factory or thick pine forest that will uniformly slow you over something like 10 or more meters. I myself would prefer a stack of pillows at least 40 meters high... but then I have been known to crack a rib just falling a meter or so playing basketball.

The amazing thing is that a number of people have been reliably documented[78] to have survived just such a fall, often with a stopping distance of only a very few meters if that, from falls as high as 18,000 feet. Sure, they usually survive with horrible injuries, but in a very few cases, e.g. falling into a deep bank of snow at a grazing angle on a hillside, or landing while strapped into an airline seat that crashed down through a thick forest canopy they haven't been particularly badly hurt...

Kids, don't try this at home! But if you ever *do* happen to fall out of an airplane at a few thousand feet, isn't it nice that your physics class helps you have the best possible chance at surviving?

Example 2.2.2: Solution to Equations of Motion for Stokes' Drag

We don't have to work very hard to actually find and solve the equations of motion for a streamlined object that falls subject to a Stokes' drag force.

We begin by writing the total force equation for an object falling down subject to near-Earth gravity and Stokes' drag, with down being positive:

$$mg - bv = m\frac{dv}{dt} \tag{2.47}$$

(where we've selected the *down* direction to be positive in this one-dimensional problem) and then ***separate variables***, reducing it to a couple of simple integrals.

[78] http://www.greenharbor.com/fffolder/ffresearch.html This website contains ongoing and constantly updated links to contemporary survivor stories as well as historical ones. It's a fun read.

That is, we isolate the velocity derivative by itself, factor out the coefficient of v on the right, divide through the v-term from the right, multiply through by dt, integrate both sides, exponentiate both sides, and rearrange. Of *course*...

Was that too fast for you[79]? Like this:

$$\begin{aligned}
\frac{dv}{dt} &= g - \frac{b}{m}v \\
\frac{dv}{dt} &= -\frac{b}{m}\left(v - \frac{mg}{b}\right) \\
\frac{dv}{v - \frac{mg}{b}} &= -\frac{b}{m}dt \\
\int_0^{v(t)} \frac{dv}{v - \frac{mg}{b}} &= -\int_0^t \frac{b}{m}dt \\
\ln\left(\frac{v(t) - \frac{mg}{b}}{-\frac{mg}{b}}\right) &= -\frac{b}{m}t \\
e^{\ln\left(\frac{v(t) - \frac{mg}{b}}{-\frac{mg}{b}}\right)} &= -e^{\frac{b}{m}t} \\
\frac{v(t) - \frac{mg}{b}}{-\frac{mg}{b}} &= -e^{\frac{b}{m}t} \\
v(t) - \frac{mg}{b} &= -\frac{mg}{b}e^{-\frac{b}{m}t} \\
v(t) &= \frac{mg}{b}\left(1 - e^{-\frac{b}{m}t}\right) \qquad (2.48)
\end{aligned}$$

or

$$\boxed{v(t) = v_t\left(1 - e^{-\frac{b}{m}t}\right)} \quad \text{with} \quad v_t = \sqrt{\frac{mg}{b}} \qquad (2.49)$$

This method uses *definite* integrals, note well, matching the (known and sought) limits of v with the corresponding limits of t. One can use indefinite integrals as well, but this leads to confusion over the dimensions, since the arguments of $\ln(x)$ and e^x must be ***dimensionless*** in physics but it is not in the context of an indefinite integral.

Objects falling through a medium under the action of *Stokes'* drag experience an *exponential* approach to a constant (terminal) velocity. This is an enormously useful piece of calculus to master; we will have a number of further opportunities to solve equations of motion *just like this* in both mechanics and electromagnetism, as linear, first order differential equations of motion have exponential solutions that are commonplace in nature – in physics, chemistry, and biology and even economics (where it is the math of compound interest and taxation). Exponential growth, exponential approach to saturation, and exponential decay are things you should understand *conceptually* as well as have an idea of how the math itself works.

Given (or rather, having solved for) $v(t)$, it isn't too difficult to integrate again and find $x(t)$, if we care to, but in this class we will usually stop here as $x(t)$ has pieces that are both linear and exponential in t and isn't as "pretty" as $v(t)$ is.

[79]Just kidding! I know you (probably) have no idea how to do this. That's why you're taking this course!

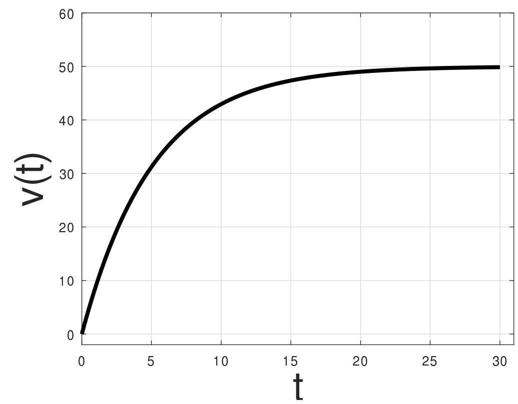


Figure 2.10: A simple object falling through a fluid experiences a drag force of $F_d = -bv$. In the figure above $m = 100$ kg, $g = 9.8$ m/sec^2, and $b = 19.6$, so that terminal velocity is 50 m/sec. Compare this figure to figure 2.8 below and note that it takes a relatively longer time to reach the same terminal velocity for an object of the same mass. Note also that the b that permits the terminal velocities to be the same is much larger than c!

2.2.4: Advanced: Solution to Equations of Motion for Turbulent Drag

Turbulent drag is set up exactly the same way that Stokes' drag is We suppose an object is dropped from rest and almost *immediately* converts to a turbulent drag force. This can easily happen because it has a bluff shape or an irregular surface together with a large coupling between that surface and the surrounding fluid (such as one might see in the following example, with a furry, fluffy ram).

The one "catch" is that the *integral* you have to do is a bit difficult for most physics students to do, unless they were really good at calculus. We will use a special method to solve this integral in the example below, one that I commend to all students when confronted by problems of this sort.

Example 2.2.3: Dropping the Ram

The UNC ram, a wooly beast of mass M_{ram} is carried by some naughty (but intellectually curious) Duke students up in a helicopter to a height H and is thrown out. On the ground below a student armed with a radar gun measures and records the velocity of the ram as it plummets toward the vat of dark blue paint below[80]. Assume that the fluffy, cute little ram experiences a turbulent drag force on the way down of $-cv^2$ in the direction shown.

In terms of these quantities (and things like g):

a) Describe *qualitatively* what you expect to see in the measurements recorded by the student ($v(t)$).

b) What is the actual algebraic solution $v(t)$ in terms of the givens.

[80] Note well: No *real* sheep are harmed in this physics problem – this actual experiment is only conducted with soft, cuddly, *stuffed* sheep...

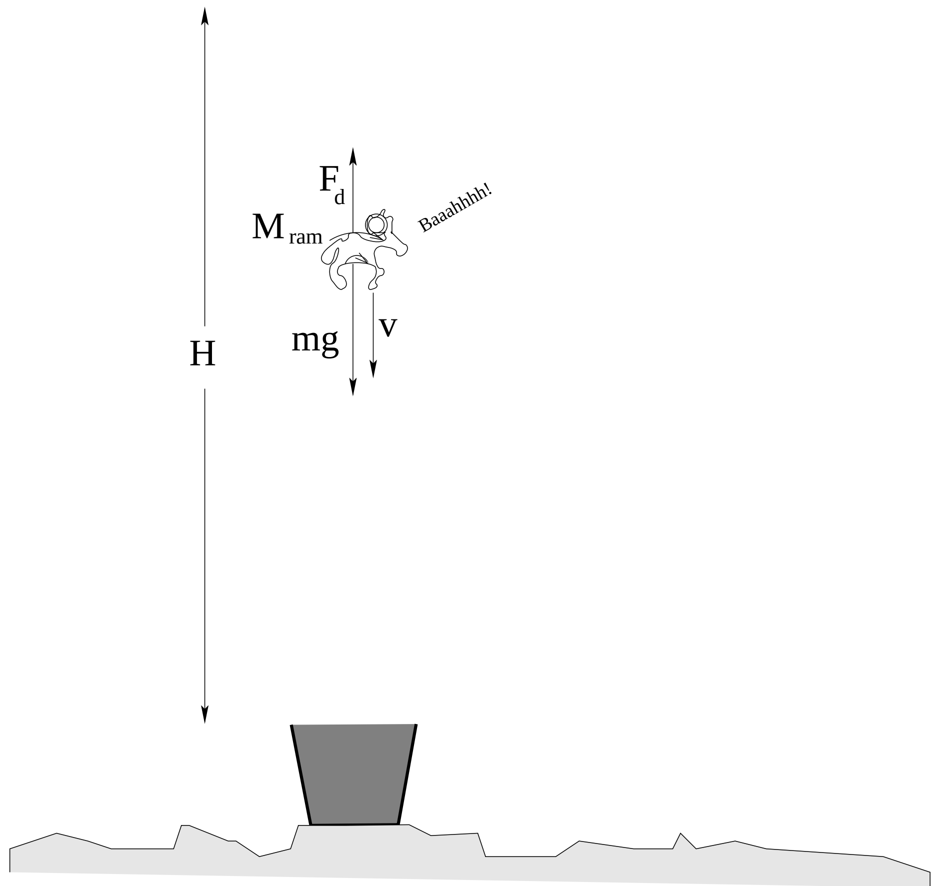


Figure 2.11: The kidnapped UNC Ram is dropped a height H from a helicopter into a vat of *Duke Blue paint!*

c) Approximately how fast is the fat, furry creature going when it splashes into the paint, more or less permanently dying it Duke Blue, if it has a mass of 100 kg and is dropped from a height of 1000 meters, given $c = 0.392$ Newton-second^2/meter^2?

Newton's Second Law for the ram is:

$$F_x = mg - cv^2 = ma = m\frac{dv}{dt} \tag{2.50}$$

and leads to the following *nonlinear* first order differential equation of motion:

$$a = \frac{dv}{dt} = g - \frac{c}{m}v^2 = -\frac{c}{m}\left(v^2 - \frac{mg}{c}\right) \tag{2.51}$$

As before, we *separate variables* to get all of the v stuff on the left, and all of the t stuff on the right, setting up a *definite* integral on both sides with corresponding limits:

$$\begin{aligned}
\frac{dv}{v^2 - \frac{mg}{c}} &= -\frac{c}{m}dt \\
\int_0^{v(t)} \frac{dv}{v^2 - \frac{mg}{c}} &= -\frac{c}{m}\int_0^t dt \\
\int_0^{v(t)} \frac{dv}{v^2 - \frac{mg}{c}} &= \int_0^{v(t)} \frac{dv}{v^2 - v_t^2} = -\frac{c}{m}t
\end{aligned} \tag{2.52}$$

where $v_t = \sqrt{mg/c}$.

Unfortunately, the remaining integral is one you aren't likely to remember. Does this mean that we are done? Not at all! There are two ways to approach it.

For most students the best way to solve an unknown integral that appears to be outside of the bounds of five integrals you should know for this class is going to be the ***look it up in an integral table*** method of solving it, also known as the ***famous mathematician method!*** Once upon a time famous mathematicians (and perhaps some not so famous ones) worked all of this sort of thing out. Heck, once upon a time *you and I* probably worked out how to solve this sort of integral in a calculus class but it isn't exactly an integral you run into every day, and I (at least) took calculus back when *Nixon* was president and my memory of integrals of this sort is a bit challenged! So what the heck, look it up!

On the other hand, if you are (or would like to be) an *advanced* student, building skills you will need in future, more complicated physics courses – perhaps a physics or math major, or at least somebody who has *fun* with stuff like this – and really *want* to see how to solve it, here it is. Students who prefer the "and then a miracle occurs to give us the result" solution may freely skip all the material between separator lines. as you will not be responsible oor doing this in any homework, quiz, or exam you might take.

Solution: This is easiest to do as a *hyperbolic trig substitution* integral. First, then, let us recall a few (relevant) hyperbolic trig identities, derivatives, and integrals (where we can often remember the latter by analogy with trig integrals):

$$\cosh u = \frac{1}{2}\left(e^u + e^{-u}\right) \qquad \sinh u = \frac{1}{2}\left(e^u - e^{-u}\right) \tag{2.53}$$

$$\cosh^2 u - \sinh^2 u = 1 \qquad \Leftrightarrow \qquad \cosh^2 - 1 = \sinh^2 u \tag{2.54}$$

$$d\cosh u = \sinh u \qquad d\sinh u = \cosh u \qquad \int \frac{du}{\sinh u} = \tanh^{-1} u + C \tag{2.55}$$

Using these, we will integrate: $\int \frac{dv}{v^2 - v_t^2}$ with a few well-chosen u-substitutions and analogy with better-known trig integrals. We'll start by letting $x = v/v_t$:

$$\int \frac{dv}{v^2 - v_t^2} = \frac{1}{v_t}\int \frac{dv/v_t}{(v/v_t)^2 - 1} = \frac{1}{v_t}\int \frac{dx}{x^2 - 1} \tag{2.56}$$

The denominator should remind us of the hyperbolic addition rule 2.54 above, so we'll try the following u-substitution: $x = v/v_t = \cosh u$ (plus a couple of items from the identities above):

$$\begin{aligned}\int \frac{dx}{x^2-1} &= \int \frac{d\cosh u}{\cosh^2 u - 1} \\ &= \int \frac{\sinh u\, du}{\sinh^2 u} \\ &= \int \frac{du}{\sinh u} \\ &= \tanh^{-1} u + C\end{aligned}$$

(where we won't need the constant of integration as we're doing definite integrals anyway).

This is the result that we might have obtained from an integral table with a certain amount of diligent search instead of by just doing it, but honestly, we likely wouldn't have gotten here much faster as we would still have had to do most of the substitutions along the way in order to put it in the form of a tabulated integral. Either way, if we substitute this result back into our original integral, we get (noting that $\tanh 0 = 0 \Rightarrow \tanh^{-1} 0 = 0$):

$$\int_0^{v(t)} \frac{dv}{v^2 - v_t^2} = \frac{1}{v_t}\tanh^{-1}\left(\frac{v(t)}{v_t}\right) = -\frac{c}{m}t \tag{2.57}$$

Finally, if we multiply both sides by $v_t = \sqrt{mg/c}$, take the $\tanh$ of both sides, and multiply both sides by v_t again, we get:

$$v(t) = \sqrt{\frac{mg}{c}}\tanh\left(\sqrt{\frac{gc}{m}}t\right) = v_t \tanh\left(\sqrt{\frac{gc}{m}}t\right) \tag{2.58}$$

This solution is plotted for you as a function of time in figure 2.12.

Clearly, solving a problem with Rayleigh/quadratic drag is a somewhat more difficult than solving one with Stokes/linear drag, but it's almost certainly a more *realistic* description of free-fall in air than what we would get using Stokes/linear drag. It's just the way nature really is, tough luck and all that. If it's any consolation to you, at least we didn't try to integrate over the *transition* between Stokes' drag and full-blown turbulent drag for the *specific* shape of a furry ram being dropped from underneath a helicopter (that no doubt has made the air it falls through initially both turbulent and beset by a substantial downdraft) and tried to compensate for any dynamic changes in c that occur as the ram, perhaps, tumbles as it falls into ever denser air!

Those are the kinds of problems *engineers* often have to solve, perhaps using computers to solve the differential equations of motion, perhaps using empirical methods based on dropping many rams in different orientations!

The moral of the story is that Real Physics$_{\rm tm}$ is often ***not terribly easy to precisely evaluate*** when we eliminate all idealizations and simplifications from a problem, but even so, it is often still easy enough to formulate the equation of motion we need to solve and *conceptually understand* the general properties of the solution. Even if we have a hard time doing all of the calculus required to find an answer to question b) above, we should all be able to understand and draw a *qualitative* picture for a) and we should really even be able to guess without any real computation at all that the ram is moving at or near terminal velocity by the time it has fallen 1000 meters.

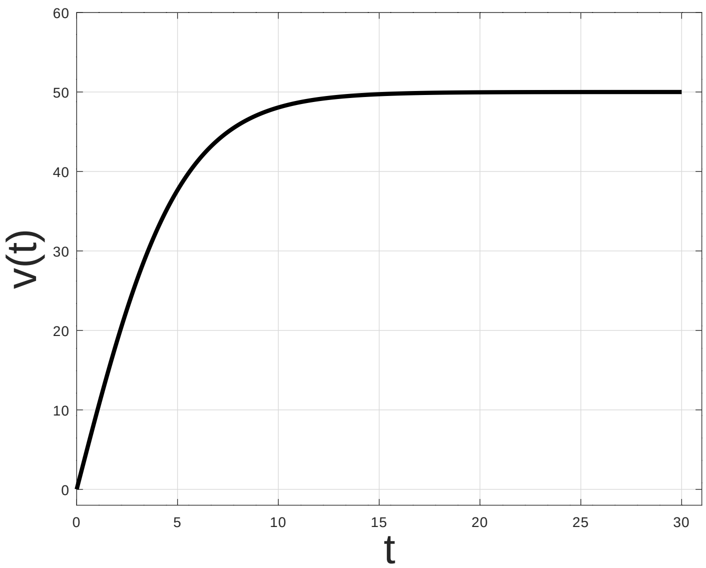


Figure 2.12: A simple object falling through a fluid experiences a drag force of $F_d = -cv^2$. In the figure above (generated using the numbers given in the ram example), $m = 100$ kg, $g = 9.8$ m/sec^2, and $c = 0.392$, so that terminal velocity is 50 m/sec. Note that the initial acceleration is g, but that after falling around 14 seconds the object is travelling at a speed very close to terminal velocity. Since even *without* drag forces it takes $\sqrt{2H/g} \approx \sqrt{200} \approx 14$ seconds to fall 1000 meters, it is almost certain that the ram will be travelling at the terminal velocity of 50 m/sec as it hits the paint!

2.3: Inertial Reference Frames

We have already spoken about coordinate systems, or "frames", that we need to imagine when we create the mental map between a physics *problem* in the abstract and the supposed *reality* that it describes. One immediate problem we face is that there are many frames we might choose to solve a problem in, but that our choice of frames isn't *completely* arbitrary. We need to reason out how much freedom we have, so that we can use that freedom to make a "good choice" and select a frame that makes the problem relatively simple.

Students that go on in physics will learn that there is more to this process than meets the eye – the symmetries of frames that preserve certain quantities actually leads us to an understanding of conserved quantities and restricts acceptable *physical theories* in certain key ways. But even students with no particular interest in relativity theory or quantum theory or advanced classical mechanics (where all of this is developed) have to understand the ideas developed in this section, simply to be able to solve problems efficiently.

2.3.1: Time

Let us start by thinking about time. Suppose that we wish to time a race (as physicists). The first thing to do is to *understand* the conditions that define the "start of the race" and the "end of the race"for any runner.

The start of the race is the instant in time that the gun goes off and the racers (as particles located at some specific point "attached" to each racer) start accelerating towards the finish line. This is a concrete definition of an actual event that you can "instantly" observe[81]. Similarly, the end of the race is the instant in time that the racers cross the finish line, where we will treat the racers again as particles and observe the time that the single specific point attached to each racer crosses this line. Again we will assume that we can observe the actual instant of each finish line event.

Consider three observers timing the same racer. One is essentially mechanical, and uses a "perfect" stop watch for each racer, one that is automatically triggered by the gun and automatically stopped by that racer crossing the finish line, effectively recording the *coincidence between two events* at the beginning and end of the race – the gun and a *simultaneous* event starting the watch, and the crossing of the finish line and a *simultaneous* event stopping that racer's watch[82]. For this observer, the race starts at time $t = 0$ *on the stop watch*, and stops at time t_f *on the same stop watch!*

The second doesn't have a stop watch – she has to use her own wristwatch set to local time. When the gun goes off she *simultaneously* observes her watch and records t_0, the time *in approximate, local time zone coordinates* her watch reads at the start of the race. When each racer crosses the finish line, she records t_1 (once for each racer, assuming she is so lightning fast and never makes mistakes), the time her watch *simultaneously* reads. To find the time of the race, she converts her watch's time to seconds and subtracts to obtain $t_f = t_1 - t_0$, which she expects to *agree* with the first observer.

The third observer has just arrived from India, and hasn't had time to reset his watch. He *also* records t'_0 for the start, t'_1 for the finish, and subtracts to once again obtain $t_f = t'_1 - t'_0$, and once again, is recording the *coincidence between distinct events* – the start or finish of the race (per racer) and the state of *something entirely different*, namely, *his presumably precisely reliable watch!*

All three of these times must agree because clearly the time required for the racer to *actually* start the race and *actually* cross the finish line has nothing to do with the observers, it has something to do with the *racer* in relation to the world. We would expect the racer to take some exact amount of time to complete the race even if there were no observers at all present or if their watches kept poor time or were read inaccurately by their observers, and it is a fundamental tenet of *classical, non-relativistic physics* that as long as these three observers *do* use watches that are perfectly accurate and those watches *are* read and recorded in perfect simultaneity with the starting and ending of each racer's race, ***all three observers must record the same duration of the race***[83]***!***

In physics we express this ***invariance principle*** by stating that we can change clocks counting time in precise SI-unit seconds at will when considering a particular problem by

[81] For the purpose of this example we will ignore things like the speed of sound or the speed of light and assume that our observation of the gun going off is instantaneous.

[82] This is the actual timing method used in e.g. swim meets where a horn electrically starts the time for all lanes and where finishing consists of a swimmer physically touching a sensor plate that stops their own lane's timer, accurate to hundredths of a second, at the end of the race.

[83] Students going on in physics should be aware that in the real, *general relativistic* Universe those times might well *not* agree, a state of affairs that makes the brains of physics students explode and have to be reassembled bit by bit to accommodate this startling fact.

means of the transformation:

$$t' = t - t_0 \tag{2.59}$$

where t_0 is the time in our *old* time-coordinate frame that we wish to be *zero* in our new, primed frame. This is basically a linear *change of variables*, a so-called "u-substitution" in calculus, but because we shift the "zero" of our clock in all cases by a *constant*, it is true that:

$$dt' = dt \tag{2.60}$$

so differentiation by t' is identical to differentiation by t and:

$$\vec{\boldsymbol{F}} = m\frac{d^2\vec{\boldsymbol{x}}}{dt^2} = m\frac{d^2\vec{\boldsymbol{x}}}{dt'^2} \tag{2.61}$$

That is, *Newton's second law* is *invariant* under uniform translations of time, so we can start our clocks whenever we wish and still accurately describe all motion relative to that time.

2.3.2: Space

We can reason the same way about space. If we want to measure the distance between two points on a line, we can do so by putting the zero on our meter stick at the first and reading off the distance of the second, or we can put the first at an arbitrary point, record the position of the second, and subtract to get the same distance. In fact, we can place the origin of our coordinate system anywhere we like and measure all of our locations *relative* to this origin, just as we can choose to start our clock at any time and measure all times relative to that time. However, this process is complicated by the fact that space is *three dimensional* and that hence particle locations are determined by a mix of *vectors* and *geometry* in any given frame.

To keep our pictures simple, we'll just visualize what's going on in *two*-dimensional frames; this is enough for us to see what's going on in spaces with more than just one dimension but not so "messy" as trying to draw 3D trajectories in a 2D projection onto a page. In figure 2.13, a single particle is moving along the trajectory indicated by the solid line, possibly speeding up and slowing down as it changes direction.

To bootstrap our discussion, we will assume that the frame S portrayed above is one where Newton's Second Law holds as long as the total force $\vec{\boldsymbol{F}}_{\rm tot}$ is made up only of *named force laws, each describing a pairwise interaction satisfying Newton's Third Law!* This latter requirement, as we will see, is crucial to our eventual definition and identification of intertial reference frames.

This means that, in S, the trajectory of the particle is described by a *vector valued trajectory function* $\vec{\boldsymbol{x}}(t)$ that is the solution to:

$$\vec{\boldsymbol{F}}_{\rm tot} = \sum_i \vec{\boldsymbol{F}}_i = m\vec{\boldsymbol{a}} \quad \Rightarrow \quad \vec{\boldsymbol{v}}(t) \quad \Rightarrow \quad \vec{\boldsymbol{x}}(t)$$

where every $\vec{\boldsymbol{F}}_i$ involves an interaction with *some other particle(s) somewhere else in the Universe!* Particles are not permitted to exert a force on *themselves*, nor can particles experience a force that is "just there" without any agent that *exerts* that force directly or indirectly!

In two more chapters, we will actually write this out in detail when we consider *many particle systems*, and you may find it useful to look ahead briefly at the chapter summary and perhaps

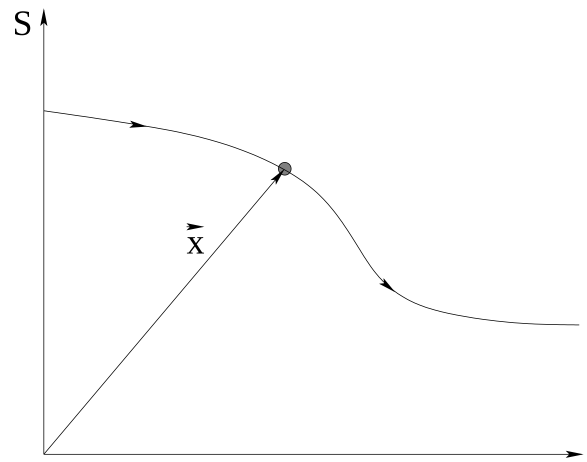


Figure 2.13: The trajectory of a single particle moving in a coordinate frame S under the influence of an arbitrary force in a frame where ***Newton's Second Law works for the total applied external force*** acting on the particle.

the first section or two there (not trying too hard yet, of course) and then come back to here so that you can see that exactly what is meant by $\vec{\boldsymbol{F}}_{\rm tot}$ being a sum of external forces exerted by masses *other* than the particle in question portrayed in 2.13.

Next, we will try to imagine this trajectory represented in a *second* coordinate frame S': As you can see in figure 2.14, if $\vec{\boldsymbol{x}}_f$ describes the vector displacement of the origin of the new

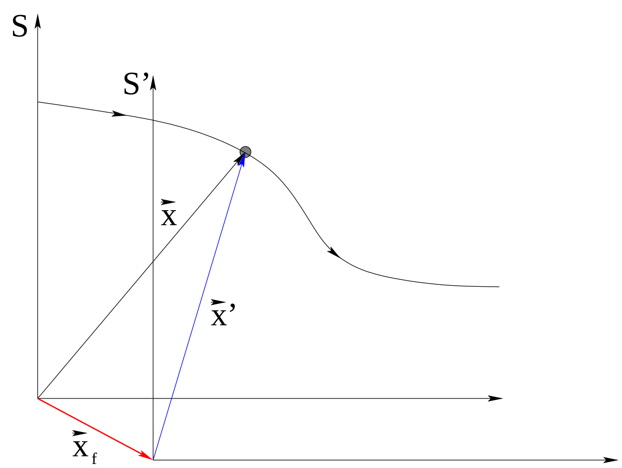


Figure 2.14: The (same) trajectory of a single particle moving in a trajectory $\vec{\boldsymbol{x}}(t)$ in coordinate frame S *and* moving in a trajectory $\vec{\boldsymbol{x}}'(t)$ relative to a *new* frame S'. The vector $\vec{\boldsymbol{x}}_f(t)$ describes the (possibly time-dependent) location of the origin of S' *relative* to the origin of S.

frame S' relative to the origin of S, we can use the *triangle rule for vector addition* to write:

$$\vec{\boldsymbol{x}} = \vec{\boldsymbol{x}}' + \vec{\boldsymbol{x}}_f \qquad \text{and} \qquad \vec{\boldsymbol{x}}' = \vec{\boldsymbol{x}} - \vec{\boldsymbol{x}}_f \tag{2.62}$$

where we can use the first to go from S' to S and the second to go from S to S' at will.

Next, we will associate time t with coordinate frame S, and t' with coordinate frame S', and use the fact that even if $t' = t - t_f$ for some shift in the origin of *time* between the two frames, as long as they are both recorded by reliable clocks counting SI-unit seconds:

$$dt = dt'$$

and we can then differentiate both sides of equation 2.62 with respect to time in *either* frame once or twice to get:

$$\vec{v} = \vec{v}' + \vec{v}_f \qquad \text{and} \qquad \vec{v}' = \vec{v} - \vec{v}_f \tag{2.63}$$

$$\vec{a} = \vec{a}' + \vec{a}_f \qquad \text{and} \qquad \vec{a}' = \vec{a} - \vec{a}_f \tag{2.64}$$

which again allow us to transform velocities or accelerations back and forth freely between descriptions in the S and S' frames.

Now let's consider Newton's Second Law again, expressed in the two frames. We agreed the S is a frame where N2 is guaranteed to be true as long as $\vec{F}_{\rm tot} = \sum_i \vec{F}_i$ where all of the $\vec{F}_i$ are named forces arising from the interaction of this particle with *other* stuff, *somewhere else*. We can thus legitimately substitute the transformation from $\vec{a}$ to $\vec{a}'$ in on the right hand side of N2, and as long as the actual physical push or pull on a particle due to another particle doesn't depend on our *description* of that push or pull in any given coordinate frame, we expect N2 to look like:

$$\vec{F}_{\rm tot} = \sum_i \vec{F}_i = m\vec{a} = m\vec{a}' + m\vec{a}_f \tag{2.65}$$

or N2 in the coordinates $\vec{x}'$ of the frame S' will look like:

$$\vec{F}_{\rm tot} = \sum_i \vec{F}_i - m\vec{a}_f = \sum_i \vec{F}_i + \vec{F}_{\rm pseudo} = m\vec{a}' \tag{2.66}$$

where we define the ***pseudoforce*** that arises from the ***acceleration of the S' frame relative to the S frame*** as:

$$\vec{F}_{\rm pseudo} = -m\vec{a}_f \tag{2.67}$$

or, the *negative* of the mass of the particle times the acceleration of the S' frame.

We call this a *pseudo* force because it isn't real! In particular, there is no object, anywhere in the Universe, interacting with the mass m by any named force law of nature, that causes this "force" to appear in N2 in the S' frame. It's just there. Worse, because there are *two distinct ways* a frame S' can accelerate relative to S – its origin can accelerate as we describe above or the frame S' can *rotate* around its own origin as a function of time, the functional description of the pseudoforces appearing in any given accelerating frame S' can be, ahem, (cough, cough) "challenging". So much so that we will treat only the very simplest of examples in this introductory course – a constant linear acceleration and some simple examples of rotating frames – and advise future physics majors that the invariance (or not) of physical laws under various coordinate transformations will form an interesting and rewarding – seriously – part of their future lives.

2.3.3: The Definition/Identification of an Inertial Reference Frame

We are now prepared to define an inertial reference frame (IRF). An inertial reference frame is any frame in which Newton's Second Law holds *without pseudoforces*. That is, as long as an *isolated* particle that isn't interacting with *any* other particles in the Universe does *not* accelerate (satisfying Newton's *First* Law) in a frame than that frame is an IRF, because if it were not, there would be a pseudoforce! Furthermore, once we have identified a *single* IRF, S, we can transform to an infinite number of alternative IRFs S' by insisting that:

$$\vec{\boldsymbol{a}}_f = 0 \quad \Rightarrow \quad \vec{\boldsymbol{v}}_f = \text{ a constant vector } \quad \Rightarrow \quad \vec{\boldsymbol{x}}_f = \vec{\boldsymbol{v}}_f t + \vec{\boldsymbol{x}}_0 \tag{2.68}$$

In this case:

$$\begin{aligned}
&\vec{\boldsymbol{x}} = \vec{\boldsymbol{x}}' + \vec{\boldsymbol{v}}_f t + \vec{\boldsymbol{x}}_0 \qquad \text{and} \qquad \vec{\boldsymbol{x}}' = \vec{\boldsymbol{x}} - \vec{\boldsymbol{v}}_f t - \vec{\boldsymbol{x}}_0 \\
&\vec{\boldsymbol{v}} = \vec{\boldsymbol{v}}' + \vec{\boldsymbol{v}}_f \qquad \text{and} \qquad \vec{\boldsymbol{v}}' = \vec{\boldsymbol{v}} - \vec{\boldsymbol{v}}_f \\
&\vec{\boldsymbol{a}} = \vec{\boldsymbol{a}}'_f \qquad \text{and} \qquad \vec{\boldsymbol{a}}' = \vec{\boldsymbol{a}} \\
&\ t = t' \qquad \text{and} \qquad t' = t
\end{aligned} \tag{2.69}$$

and

$$\vec{\boldsymbol{F}}_{\rm tot} = \sum_i \vec{\boldsymbol{F}}_i = m\vec{\boldsymbol{a}} = m\vec{\boldsymbol{a}}' \tag{2.70}$$

and N2 holds in both S and S'.

Any frame moving at a constant velocity relative to an IRF is also an IRF! Also, the freedom to change origins of our coordinate frame corresponds precisely to our *kinematic* freedom to add $\vec{\boldsymbol{v}}_f t + \vec{\boldsymbol{x}}_0$ to any solution to N2 in an IRF and still have a solution to N2!

Now, you might argue that it is impossible to arrange any particle in the Universe in such a way that it isn't interacting with any other particles if only because *gravitation* (as we will see in the last chapter of this textbook) is a force with infinite range that exists between every pair of particles in the Universe, so we could only have zero force in a Universe with a single particle, and I wouldn't disagree, but nevertheless we can easily arrange environments where either:

a) External forces caused by those objects interacting with the particle in question are negligibly small, perhaps because on average they *cancel* or because they are due to objects that are *very far away* with interactions that drop off rapidly with distance of separation; or

b) We can tally up the most important *named* forces of interaction, each with a partner and obeying Newton's Third Law, for the given particle due to "nearby" partners to high accuracy and can then neglect those due to objects much farther away or otherwise average them out of the picture as in a); and note that

c) Even if there *is* a leftover pseudoforce (as there is in any laboratory frame located in an actual classroom sitting on the actual surface of our rotating planet as it revolves, acceleratig,, tugged this way and that by the sun, the moon, and to a lesser extent by the other nearby planets while the sun in turn orbits the local gravitational center of the galaxy) it can easily be too small to be detectable in most lab-scale experiments!

The set of transformations above in equations 2.69 are referred to as the ***Galilean transformation between coordinate frames*** and form the basis of ***Galilean relativity***. This is the complete set of coordinate transformations that completely preserve our dynamical principle, Newton's Second Law, for named interaction forces that do not themselves depend on tranformations between IRFs and all have N3 partners. They will work extremely well in this first treatment of classical physics through gravitation, but they ultimately *fail* to work for electromagnetic forces we will study after gravitation. In order to obtain a consistent description of classical electrodynamics, we will have to deduce or derive a new, *special* relativity, because it will turn out that "magnetic" forces, in particular, depend intrinsically on particle velocity and hence are in fact *not* invariant under the underlying Galilean coordinate transformation!

To summarize, any coordinate frame travelling at a *constant velocity* (in which Newton's *first* law will thus apparently hold without any pseudoforces[84]) is an IRF, and since our law of dynamics is *invariant* with respect to changes of IRF (as long as the force laws themselves are), we have complete freedom to choose the one that is the most convenient.

Because it is sometimes useful to *change reference frames* in the middle of a problem solution in order to obtain the simplest possible answer with the truly least amount of work, and because it is *always* useful to choose the "best" reference frame for solving any given type of problem, we will now spend a bit of time trying to learn how to change reference frames from a "lab" IRF S, presumed to be "at rest" and not accelerating, into a moving frame S' that may or may not be accelerating.

2.3.4: Changing Reference Frames

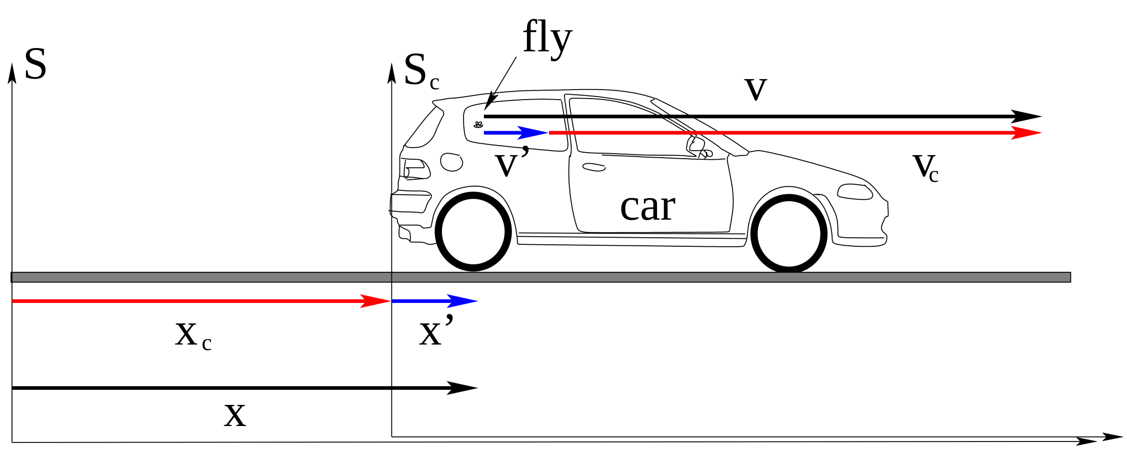


Figure 2.15: Two frames used to describe the actual motion of a fly buzzing around a car. This example can help you figure out when to *add* or *subtract* the velocity/position of two frames to change coordinates in between them.

Suppose you are driving a car at a uniform velocity down the highway (basically, 1D mo-

[84]This is a rather subtle point, as my colleague Ronen Plesser pointed out to me. If velocity itself is always defined relative to and measured within some frame, then "constant velocity" relative to *what* frame? The Universe doesn't come with a neatly labelled Universal inertial reference frame – or perhaps it does, the frame where the blackbody background radiation leftover from the big bang is isotropic – but even if it does the answer is "relative to another inertial reference frame" which begs the question, a very bad thing to do when constructing a consistent physical theory. To avoid this, an inertial reference frame may be defined to be "any frame where Newton's *First* Law is true, that is, a frame where objects at rest remain at rest and objects in motion remain in uniform motion unless acted on by an unbalanced total force that is the sum of actual named external pairwise forces", as it is above, a specification that does not require that we identify an IRF relative to some *special* IRF.

tion). Inside the car is a fly, flying from the back of the car to the front of the car, as portrayed in figure 2.15. To *you*, the fly is moving slowly, as represented by the (blue) vector v' in the figure, obtained by differentiating its (blue) position vector x' relative to an origin at the rear bumper of the car that defines the frame S_c of the car.

To an observer on the *ground*, however, the car's rear bumper is itself at the (red) position vector x_c and moving away from the ground origin in frame S at (red) velocity vector v_c. To find the position and velocity of the *fly* relative to the ground, it is pretty obvious that at the instant portrayed, the (black) vector position and velocity:

$$x = x' + x_c = \quad \text{and} \quad v = v' + v_c \tag{2.71}$$

(red plus blue) tell us how to find the position or velocity of the fly relative to the ground given its position and velocity relative to the (1D) car as well as the position and velocity of the car.

It's as simple as that! If you can remember this one example, you can always figure out whether you should add or subtract velocity to move into or out of a given reference frame! Therefore, repeat the following ritual expression (and meditate) until it makes sense *forwards and backwards*:

> The position of the fly in the coordinate frame of the ground is the position of the fly in the coordinate frame of the car, plus the position of the car in the coordinate frame of the ground.

In this way, even if the fly isn't moving in a straight line but is buzzing back and forth and sideways inside the car, as long as the car is an *inertial reference frame* travelling at a *constant velocity relative to the ground* we can find the *for real* vector position of the fly in time in either frame give its vector position in the other as:

$$\begin{aligned} \vec{x}(t) &= \vec{x}'(t) + \vec{v}_{\text{frame}} t \quad (+\vec{x}_0) \qquad \text{or} \\ \vec{x}'(t) &= \vec{x}(t) - \vec{v}_{\text{frame}} t \quad (-\vec{x}_0) \end{aligned} \tag{2.72}$$

where once again these equations (and their derivatives, and the assumption that $t = t'$) are called the ***Galilean transformation***.

Note that we we could start our clock at a time when $\vec{x}(t = 0) = \vec{x}_0$ is the displacement of the origin of the "moving" frame, hence the additional term in parentheses. We will often choose to make S and S' have the *same origin* at time $t = 0$ to eliminate the need for this term.

The physics of the fly relative to (expressed in) the coordinate frame in the car are identical to the physics of the fly relative to (expressed in) the coordinate frame on the ground when we account for all of the physical forces (in either frame) that act on the fly. Indeed, as any child knows, you can play catch with a sibling sharing the seat with you in a car moving at constant velocity and the ball behaves exactly as it does if you are playing catch on a sofa "at rest" in your living room[85]. Anyone who has ridden in a subway has probably experienced the strangeness of looking through the windows at a train on the next track over as it starts to move, and having to suddenly look our of windows on the other side to see if it is moving or if

[85]...which is actually zooming along at maybe 1000 kilometers/hour relative to the axis of the rotating earth, as the earth itself zips around the sun at roughly 30 kilometers per *second*...

it is your *own* car that is moving *the other way*. If the acceleration of the car is gentle enough that the pseudoforces are negligible, it is not at all easy to tell!

The Galilean transformation isn't the only possible way to relate frames, and in fact it doesn't correctly describe nature. A different transformation called the *Lorentz* transformation from the *theory of relativity* works much better, where both length intervals and time intervals *change* when changing inertial reference frames. However, describing and deriving relativistic transformations (and the postulates that lead us to consider them in the first place) is beyond the scope of this course, and they are not terribly important in the classical regime where things move at speeds much less than that of light.

2.4: Non-Inertial Reference Frames – Pseudoforces

Let's return to the case where S is indeed an IRF, but where S' is *not*, where (for example) the S' frame is *uniformly accelerating* relative to the frame S. We found that N2 in the accelerating non-IRF S' must be written as:

$$\vec{F}'_{\rm tot} = \vec{F}_{\rm tot} - m\vec{a}_{\rm frame} = \vec{F}_{\rm tot} - \vec{F}_p = m\vec{a}' \tag{2.73}$$

where $\vec{F}_{\rm tot}$ is the sum of all *real, interaction based forces* acting on the particle and $\vec{F}_p$ is a ***pseudoforce*** – a force that does not exist as any combination of *interaction* forces of nature and that hence has *no Newton's Third Law partner*. It isn't a force at all – it is a ***kinematic, not dynamic*** term that arises from the math/geometry/calculus of the frame transformation but that we must move over to the force side of the N2 equation if we wish to solve something that *looks* like N2 but works in the accelerating frame!

In the case of uniform frame accelerations, this pseudoforce is proportional to the mass times a the constant acceleration of the frame and behaves a lot like the only force rule we have so far which produces uniform forces proportional to the mass – gravity near Earth's surface! Indeed, it feels to our senses like gravity has been *modified* if we ride along in an accelerating frame – made weaker, stronger, changing its direction. However, our algebra above shows that a pseudoforce behaves *consistently* like that – we can actually solve equations of motion in the accelerating frame using the additional "force rule" of the pseudoforce and we'll *get the right answers* within the frame and, when we add the coordinates in the frame to the ground/inertial frame coordinates of the frame, in those coordinates as well.

Pseudoforces are forces which aren't really there. Why, then, you might well ask, do we deal with them? From the previous paragraph you should be able to see the answer: because it is psychologically and occasionally computationally useful to do so. Psychologically because they describe what we experience in such a frame; computationally because we *live* in a non-inertial frame (the surface of the rotating earth) and for certain problems it is the solution in the natural coordinates *of* this non-inertial frame that matters.

We have encountered a few pseudoforces already, either in the course or in our life experience. We will encounter more in the weeks to come. Here is a short list of places where one experiences pseudoforces, or might find the concept itself useful in the mathematical description of motion in an accelerating frame:

a) The force added or subtracted to a real force (i.e. – mg, or a normal force) in a frame accelerating uniformly. The elevator and boxcar examples below illustrate this nearly ubiquitous experience. This is the "force" that pushes you back in your seat when riding in a jet as it takes off, or a car that is speeding up. Note that it *isn't* a force at all – all that is *really* happening is that the seat of the car is exerting a normal force on you so that you accelerate at the same rate as the car, but this *feels like gravity has changed* to you, with a new component added to mg straight down.

b) Rotating frames account for lots of pseudoforces, in part because we live on one (the Earth). In a frame attached to a particle moving in a circle of radius r at constant speed v, if we moving the mass times the *kinematic* acceleration v^2/r to the other side of N2, it becomes a pseudoforce known popularly as "centrifugal" force – a kind of weird "gravity" that throws you to the *outside* of the circle of motion. In a ball being whirled in a circle, this "centrifugal force" is e.g. equal to the tension T in a string that provides the *real centripetal* force acting on the ball. This pseudo "gravity", however, is nothing like *real* gravity, be it near the Earth or far away – it varies like $\Omega^2 r$ with r the distance from the axis of rotation as opposed to the $1/r^2$ form with r the distance from the center of the Earth) for real gravity – and has other peculiarities as well!

c) Because of this *geometrically* distinct r dependence, there are slightly different pseudoforces acting on objects falling towards or away from the center of a rotating sphere (such as the earth). A particle falling down is travelling *faster* to the East than the point on the ground directly below, and hence appears to deflect *spinward* (to the East) as it falls. A particle thrown directly up, away from the ground, is travelling more slowly than points that remain directly "above" it as the Earth's frame rotates, and appears to deflect *antispinward* (West) as it rises. Suitably formulated kinematic pseudoforcess describe this apparent deflection as a particle rises or falls.

d) Objects moving north or south along the *surface* of a rotating sphere experience a *similar* deflection, for similar reasons. As a particle moves towards the equator, it is suddenly travelling too slowly for its new radius at a constant Ω, and is apparently "deflected" antispinward/west. As it travels away from the equator it is suddenly traveling too fast for its new radius and is apparently deflected spinward/east.

These latter two pseudoforce effects in combination are referred to as the ***Coriolis*** (pseudo)***Force*** and are a major driving factor in the time evolution of weather patterns, especially spectacular storms such as hurricanes, on the rotating approximately spherical frame tied to the surface of the Earth in terms of which we perforce *observe* those trajectories. They also complicate naval artillery trajectories,, missile launches, and other long range ballistic trajectories *in* the rotating frame, as the coriolis forces combine with drag forces to produce very real and somewhat unpredictable deflections of ballistic trajectories through the atmosphere compared to firing right at a target in a presumed cartesian, flat, inertial frame.

In spite of the fact that one *can* always choose to solve physics problems in an IRF – and indeed, this is nearly always the best choice – there are times when we pretty much *have* to obtain the final solution in a rotating frame, for example describing weather where we actually live. Describing the motion of every bit of mass in a precessing bicycle wheel is vastly simpler if we do so *relative* to the rotating frame instead of relative to a point "at rest" out in space somewhere. It is similarly fairly easy to describe the oscillations of a pendulum in the frame of an accelerating elevator and add the position of the elevator to that of the pendulum *in* the elevator to get the pendulum position relative to the ground, but insanely difficult to formulate and solve and express the answer relative to the rest frame of the ground.

We will therefore consider a few example cases of pseudoforces in frames where we can see things *both ways* to get a better idea of how they work.

Example 2.4.1: Weight in a Rocket Ship

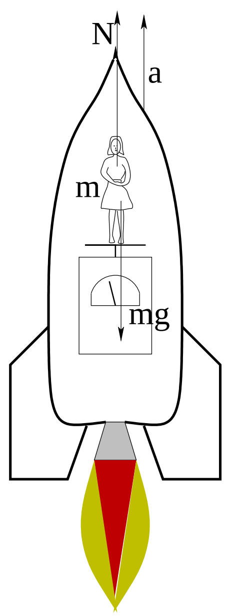


Figure 2.16: An rocket ship accelerates up with a net acceleration of a. The normal force exerted by the (scale on the floor) of the elevator on a person overcomes the force of near-Earth gravity to provide this acceleration.

Here's a 1D example. Let's compute our apparent weight in an rocket that is accelerating up (or down, but say up) at some net rate a. If you are riding in the rocket, you *must* be accelerating up with the *same* acceleration. Therefore the net force on you *must* be

$$F_{\rm tot} = \sum_i F_i = ma \;(+\mathrm{up}) \tag{2.74}$$

That net force is made up of two "real" forces: The force of gravity which pulls you "down" (in whatever coordinate frame you choose), and the (normal) force exerted by all the molecules in the scale upon the soles of your feet. This latter force is what the *scale* reads as your "weight"[86].

Thus:

$$\sum F = N - mg = ma \tag{2.75}$$

[86]Mechanically, a non-digital bathroom scale reads the net force applied to/by its top surface as that force e.g. compresses a spring, which in turn causes a little geared needle to spin around a dial. This will make more sense later, as we come study springs in more detail.

or

$$N = W = mg + ma \tag{2.76}$$

where W is your "weight" *as measured by the scale in the rocket ship!* The nerves in your feet and legs report that you have become somewhat "heavier" and the scale you stand on agrees! Note, however, that your *real* weight is still mg, as the ***interaction between you and the Earth has not changed!***

However we *could* have formulated this problem *in the frame of the rocket ship* as:

$$\begin{aligned} F_{\text{tot}} - F_p &= ma' = 0 \ (!) \\ (N - mg) - ma &= 0 \\ N = mg + ma &= m(g + a) = mg' \end{aligned} \tag{2.77}$$

because (!) your acceleration $a' = 0$ *in the accelerating frame of the rocket ship*. $W = m(g+a)$ is *still* your apparent weight in the accelerating rocket ship, but now we got it by adding a pseudoforce $-ma$ to the two real forces acting on you as the ship lifts off. Because this force scales linearly with the mass *just like near-Earth gravitation* to your legs, and the scale, it feels exactly like *gravity in the accelerating ship got stronger*, shifting to a new value of:

$$g' = g + a \tag{2.78}$$

Note that if the rocket engine then cuts off, $a \to -g$ (the rocket and all of its contents accelerate down due to near-Earth gravity, even if the ship continues going up at first). In that case the sum of gravity and pseudogravity combine as $g' = g - g = 0$, and you *feel* "weightless" in the frame of the freely falling rocket! Of course your real weight has still not changed and is still mg, as you will doubtless eventually discover if the rocket has not managed to either achieve escape speed or enter an orbit in the meantime – when it collides once again with the ground!

Example 2.4.2: Pendulum in a Boxcar

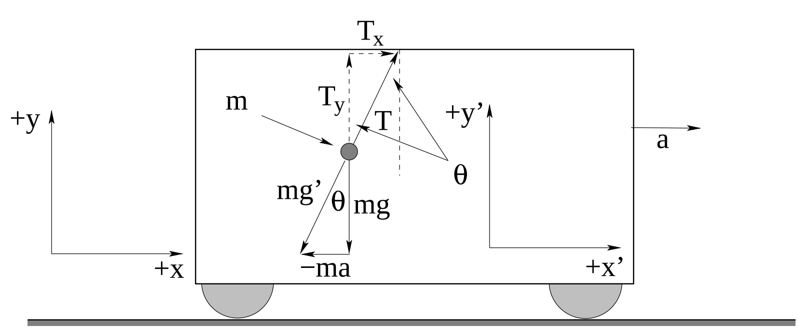


Figure 2.17: A plumb bob or pendulum hangs "at rest" at an angle θ in the frame of a boxcar that is uniformly accelerating to the right.

In figure 2.17, we see a railroad boxcar that is (we imagine) being uniformly and continuously accelerated to the right at some constant acceleration $\vec{a} = a_x\hat{x}$ in the (ground, inertial) coordinate frame shown. A pendulum of mass m has been suspended "at rest" (in the accelerating frame of the boxcar) at a stationary angle θ relative to the inertial frame y axis as shown

We would like to be able to answer questions such as:

a) What is the tension T in the string suspending the mass m?

b) What is the angle θ in terms of the givens and knows?

We can solve this problem and answer these questions *two ways* (in two distinct frames). The first, and I would argue "right" way, is to solve the Newton's Second Law *dynamics* problem in the inertial coordinate system corresponding to the ground. This solution (as we will see) is simple enough to obtain, but it does make it relatively difficult to relate the answer in ground coordinates (that isn't going to be terribly simple) to the *extremely simple* solution in the primed coordinate system of the accelerating boxcar shown in figure 2.17. Alternatively, we can solve and answer it directly in the primed accelerating frame – the coordinates you would *naturally* use if you were *riding* in the boxcar – by means of a pseudoforce.

Let's proceed the first way. In this approach, we as usual decompose the tension in the string in terms of the ground coordinate system:

$$\sum F_x = T_x = T\sin(\theta) = ma_x \tag{2.79}$$

$$\sum F_y = T_y - mg = T\cos(\theta) - mg = ma_y = 0 \tag{2.80}$$

where we see that a_y is 0 because the mass is "at rest" in y as the whole boxcar frame moves only in the x-direction and hence has no y velocity or net acceleration.

From the second equation we get:

$$T = \frac{mg}{\cos(\theta)} \tag{2.81}$$

and if we substitute this for T into the first equation (eliminating T) we get:

$$mg\tan(\theta) = ma_x \tag{2.82}$$

or

$$\tan(\theta) = \frac{a_x}{g} \tag{2.83}$$

We thus know that $\theta = \tan^{-1}(a_x/g)$ and we've answered the second question above. To answer the first, we look at the right triangle that makes up the vector force of the tension (also from Newton's Laws written componentwise above):

$$T_x = ma_x \tag{2.84}$$

$$T_y = mg \tag{2.85}$$

and find:

$$T = \sqrt{T_x^2 + T_y^2} = m\sqrt{a_x^2 + g^2} = mg' \tag{2.86}$$

where $g' = \sqrt{a_x^2 + g^2}$ is the *effective* gravitation that determines the tension in the string, an idea that won't be completely clear yet. At any rate, we've answered both questions.

To make it clear, let's answer them both again, this time using a pseudoforce in the accelerating frame of the boxcar. In the boxcar, according to the work we did above, we expect to have a total *effective* force:

$$\vec{F}' = \vec{F} - m\vec{a}_{\rm frame} \tag{2.87}$$

where $\vec{F}$ is the sum of the actual force laws and rules in the inertial/ground frame and $-m\vec{a}_{\rm frame}$ is the pseudoforce associated with the acceleration of the frame of the boxcar. In this particular problem this becomes:

$$-mg'\hat{y}' = -mg\hat{y} - ma_x\hat{x} \tag{2.88}$$

or the magnitude of the *effective* gravity in the boxcar is mg', and it points "down" in the boxcar frame in the $\hat{y}'$ direction. Finding g' from its components is now straightforward:

$$g' = \sqrt{a_x^2 + g^2} \tag{2.89}$$

as before and the direction of $\hat{y}'$ is now inclined at the angle

$$\theta = \tan^{-1}\left(\frac{a_x}{g}\right) \tag{2.90}$$

also as before. Now we get T directly from the one dimensional *statics* problem along the $\hat{y}'$ direction:

$$T - mg' = ma_{y'} = 0 \tag{2.91}$$

or

$$T = mg' = m\sqrt{a_x^2 + g^2} \tag{2.92}$$

as – naturally – before. We get the same answer either way, and there isn't *much* difference in the work required. I personally prefer to think of the problem, and solve it, in the inertial ground frame, but what you *experience* riding along in the boxcar is much closer to what the second approach yields – gravity appears to have gotten stronger and to be pointing back at an angle as the boxcar accelerates, which is *exactly what one feels* standing up in a bus or train as it starts to move, in a car as it rounds a curve, in a jet as it accelerates down the runway during takeoff.

Sometimes (rarely, in my opinion) it is convenient to solve problems (or gain a bit of insight into behavior) using pseudoforces in an accelerating frame (and the latter is certainly in better agreement with our experience in those frames) but it will lead us to make silly and incorrect statements and get problems *wrong* if we do things carelessly, such as call mv^2/r a *force* where it is really just ma_c, the *right* and side of Newton's Second Law where the left hand side is made up of actual force rules. In this kind of problem and many others it is better to just use the real forces in an inertial reference frame, and we will fairly religiously stick to this in this textbook. As the next discussion (intended only for more advanced or intellectually curious students who want to be guided on a nifty wikiromp of sorts) suggests, however, there *is* some advantage to thinking more *globally* about the apparent equivalence between gravity in particular and pseudoforces in accelerating frames.

2.4.1: Advanced: General Relativity and Accelerating Frames

As serious students of physics and mathematics will one day learn, Einstein's Theory of Special Relativity[87] and the associated Lorentz Transformation[88] will one day replace the theory of inertial "relativity" and the Galilean transformation between inertial reference frames we deduced in week 1. Einstein's result is based on more or less the same general idea – the laws of physics need to be invariant under inertial frame transformation. The problem is that *Maxwell's Equations* (as you will learn in detail in part 2 of this course, if you continue) are the actual laws of nature that describe electromagnetism and hence need to be so invariant. Since Maxwell's equations predict the speed of light, the speed of light *has to be the same in all reference frames!*

This has the consequence – which we will not cover in any sort of detail at this time – of causing space and time to become a system of *four* dimensional spacetime, not three space dimension plus time as an independent variable. Frame transformations nonlinearly mix space coordinates and time as a coordinate instead of just making simple linear tranformations of space coordinates according to "Galilean relativity".

Spurred by his success, Einstein attempted to describe force itself in terms of curvature of spacetime, working especially on the ubiquitous force of gravity. The idea there is that the pseudoforce produced by the acceleration of a frame is *indistinguishible* from a gravitational force, and that a generalized frame transformation (describing acceleration in terms of curvature of spacetime) should be able to explain both.

This isn't *quite* true, however. A uniformly accelerating frame can match the *local magnitude* of a gravitational force, but gravitational fields have (as we will learn) a global geometry that cannot be matched by a uniform acceleration – this hypothesis "works" only in small volumes of space where gravity is approximately uniform, for example in the elevator or train above. Nor can one match it with a rotating frame as the geometric form of the coriolis force that arises in a rotating frame does not match the $1/r^2\hat{\boldsymbol{r}}$ gravitational force law.

The consequence of this "problem" is that it is considerably more difficult to derive the theory of general relativity than it is the theory of special relativity – one has to work with *manifolds*[89] . In a sufficiently small volume Einstein's hypothesis is valid and gives excellent results that predict sometimes startling but experimentally verified deviations from classical expectations (such as the precession of the perihelion of Mercury)[90]

The one remaining problem with general relativity – also beyond the scope of this textbook – is its fundamental, deep incompatibility with quantum theory. Einstein wanted to view *all* forces of nature as being connected to spacetime curvature, but quantum mechanics provides

[87]Wikipedia: http://www.wikipedia.org/wiki/Special Relativity.

[88]Wikipedia: http://www.wikipedia.org/wiki/Lorentz Transformation.

[89]Wikipedia: http://www.wikipedia.org/wiki/Manifold. A manifold is a topological curved space that is locally "flat" in a sufficiently small volume. For example, using a simple cartesian map to navigate on the surface of the "flat" Earth is quite accurate up to distances of order 10 kilometers, but increasingly inaccurate for distances of order 100 kilometers, where the fact that the Earth's surface is really a curved spherical surface and not a flat plane begins to matter. Calculus on curved spaces is typically defined in terms of a manifold that covers the space with locally Euclidean patches. Suddenly the mathematics has departed from the relatively simple calculus and geometry we use in this book to something rather *difficult*...

[90]Wikipedia: http://www.wikipedia.org/wiki/Tests of general relativity. This is one of several "famous" tests of the theory of general relativity, which is generally accepted as being *almost* correct, or rather, correct in context.

a spectacularly different picture of the cause of interaction forces – the exchange of quantized particles that mediate the field and force, e.g. photons, gluons, heavy vector bosons, and by extension – gravitons[91] . So far, nobody has found an entirely successful way of unifying these two rather distinct viewpoints, although there are a number of candidates[92] .

[91] Wikipedia: http://www.wikipedia.org/wiki/gravitons. The quantum particle associated with the gravitational field.

[92] Wikipedia: http://www.wikipedia.org/wiki/quantum gravity. Perhaps the best known of these is "string theory", but as this article indicates, there are a number of others, and until dark matter and dark energy are better understood as apparent *modifiers* of gravitational force we may not be able to experimentally choose between them.

2.5: Just For Fun: Hurricanes

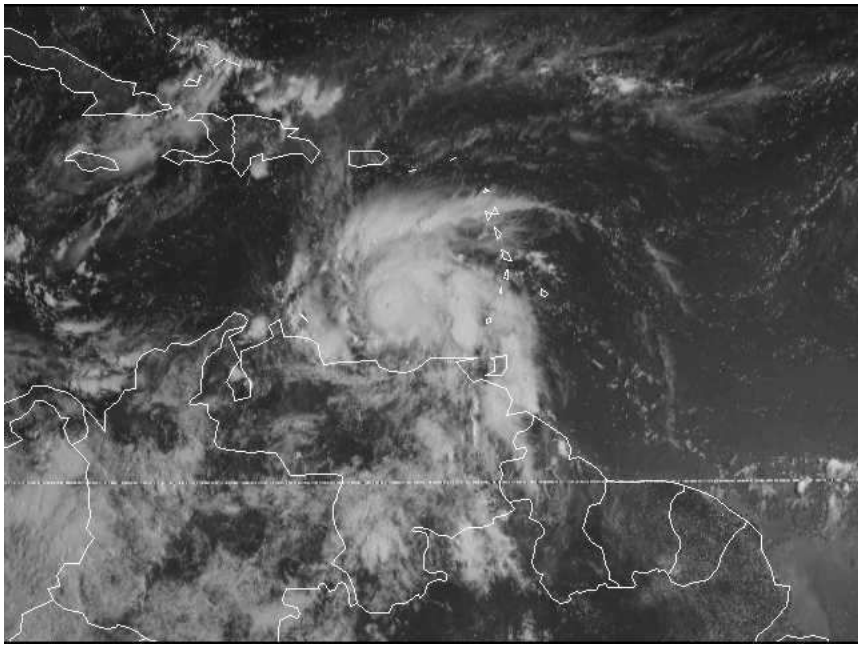


Figure 2.18: Satellite photo of Hurricane Ivan as of September 8, 2004. Note the roughly symmetric rain bands circulating in towards the center and the small but clearly defined "eye".

Hurricanes are of great interest, at least in the Southeast United States where every fall several of them (on average) make landfall somewhere on the Atlantic or Gulf coast between Texas and North Carolina. Since they not infrequently do billions of dollars worth of damage and kill dozens of people (usually drowned due to flooding) it is worth taking a second to look over their Coriolis dynamics.

In the northern hemisphere, air circulates around high pressure centers in a generally clockwise direction as cool dry air "falls" out of them in all directions, deflecting west as it flows out south and east as it flow out north.

Air circulates around low pressure centers in a counterclockwise direction as air rushes to the center, warms, and lifts. Here the eastward deflection of north-travelling air *meets* the west deflection of south-travelling air and creates a whirlpool spinning opposite to the far curvature of the incoming air (often flowing in from a circulation pattern around a neighboring high pressure center).

If this circulation occurs over warm ocean water it picks up considerable water vapor and heat. The warm, wet air cools as it lifts in the central pattern of the low and precipitation occurs, releasing the latent heat of fusion into the rapidly expanding air as wind flowing *out* of the low pressure center at high altitude in the usual clockwise direction (the "outflow" of the storm). If the low remains over warm ocean water and no "shear" winds blow at high altitude across the developing eye and interfere with the outflow, a stable pattern in the storm emerges that gradually amplifies into a hurricane with a well defined "eye" where the air has very low pressure and no wind at all.

Figure 2.21 shows a "snapshot" of the high and low pressure centers over much of North

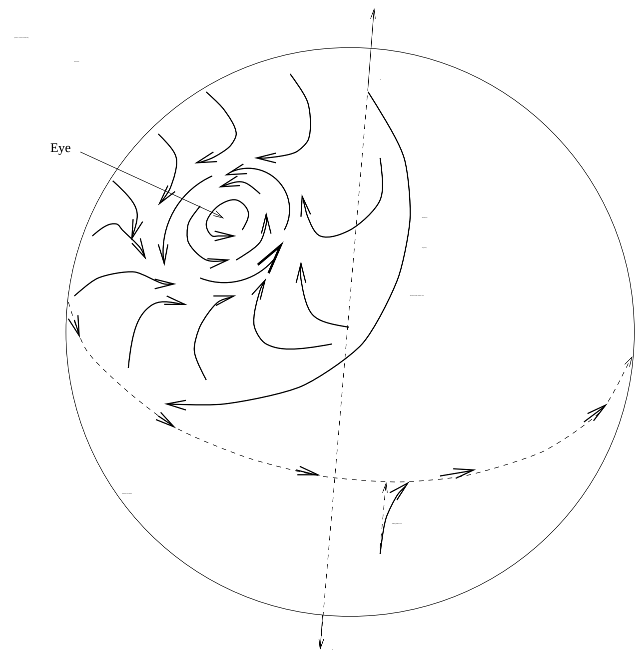


Figure 2.19

Figure 2.20: Coriolis dynamics associated with tropical storms. Air circulating clockwise (from surrounding higher pressure regions) meets at a center of low pressure and forms a counter-clockwise "eye".

and South America and the Atlantic on September 8, 2004. In it, two "extreme low" pressure centers are clearly visible that are either hurricanes or hurricane remnants. Note well the *counterclockwise* circulation around these lows. Two large high pressure regions are also clearly visible, with air circulating around them (irregularly) clockwise. This rotation smoothly transitions into the rotation around the lows across boundary regions.

As you can see the dynamics of all of this are rather complicated – air cannot just "flow" on the surface of the Earth – it has to flow *from one place to another*, being replaced as it flows. As it flows north and south, east and west, up and down, pseudoforces associated with the Earth's rotation join the *real* forces of gravitation, air pressure differences, buoyancy associated with differential heating and cooling due to insolation, radiation losses, conduction and convection, and moisture accumulation and release, and more. Atmospheric modelling is difficult and not terribly skilled (predictive) beyond around a week or at most two, at which point small fluctuations in the initial conditions often grow to unexpectedly dominate global weather patterns, the so-called "butterfly effect"[93] .

In the specific case of hurricanes (that do a lot of damage, providing a lot of political and economic incentive to improve the predictive models) the details of the dynamics and energy

[93]Wikipedia: http://www.wikipedia.org/wiki/Butterfly Effect. So named because "The flap of a butterfly's wings in Brazil causes a hurricane in the U.S. some months later." This latter sort of statement isn't really correct, of course – *many* things conspire to cause the hurricane. It is intended to reflect the fact that weather systems exhibit deterministic chaotic dynamics – infinite sensitivity to initial conditions so that tiny differences in initial state lead to radically different states later on.

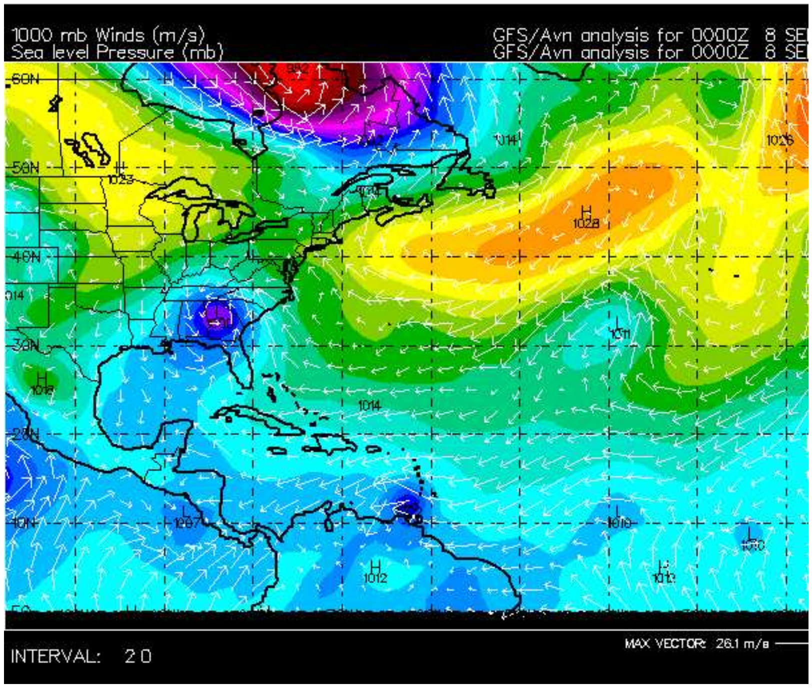


Figure 2.21: Pressure/windfield of the Atlantic on September 8, 2004. Two tropical storms are visible – the remnants of Hurricane Frances poised over the U.S. Southeast, and Hurricane Ivan just north of South America. Two more low pressure "tropical waves" are visible between South America and Africa – either or both could develop into tropical storms if shear and water temperature are favorable. The low pressure system in the middle of the Atlantic is extratropical and very unlikely to develop into a proper tropical storm.

release are only gradually being understood by virtue of intense study, and at this point the hurricane models *are* quite good at predicting motion and consequence within reasonable error bars up to five or six days in advance. There is a wealth of information available on the Internet[94] to any who wish to learn more. An article[95] on the Atlantic Oceanographic and Meteorological Laboratory's Hurricane Frequently Asked Questions[96] website contains a lovely description of the structure of the eye and the inflowing rain bands.

Atlantic hurricanes *usually* move from Southeast to Northwest in the Atlantic North of the equator until they hook away to the North or Northeast. Often they sweep away into the North Atlantic to die as mere extratropical storms without ever touching land. When they do come ashore, though, they can pack winds well over a hundred miles an hour. This is faster than the "terminal velocity" associated with atmospheric drag and thereby they are powerful enough to lift a human or a cow right off their feet, or a house right off its foundations. In addition, even mere "tropical storms" (which typically have winds in the range where wind per se does relatively little damage) can drop a foot of rain in a matter of hours across tens of thousands of square miles or spin down local tornadoes with high and damaging winds. Massive flooding, not wind, is the most common cause of loss of life in hurricanes and other tropical storms.

Hurricanes also can form in the Gulf of Mexico, the Carribean, or even the waters of the

[94] Wikipedia: http://www.wikipedia.org/wiki/Tropical Cyclone.
[95] http://www.aoml.noaa.gov/hrd/tcfaq/A11.html
[96] http://www.aoml.noaa.gov/hrd/weather_sub/faq.html

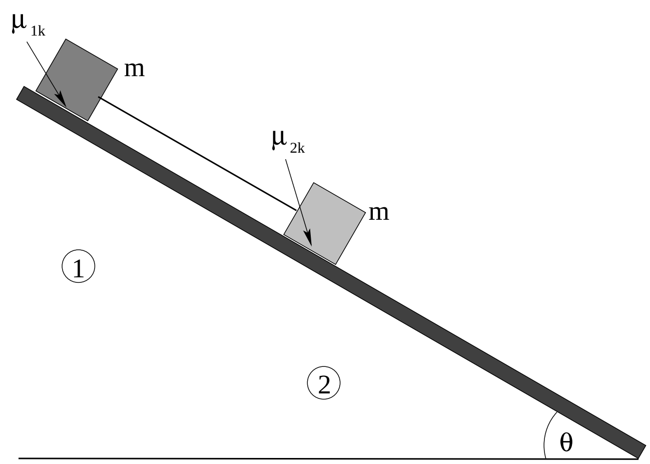
Pacific close to Mexico. Tropical cyclones in general occur in all of the world's tropical oceans *except* for the Atlantic south of the equator, with the highest density of occurrence in the Western Pacific (where they are usually called "typhoons" instead of "hurricanes"). All hurricanes tend to be highly unpredictable in their behavior as they bounce around between and around surrounding air pressure ridges and troughs like a pinball in a pinball machine, and even the best of computational models, updated regularly as the hurricane evolves, often err by over 100 kilometers over the course of just a day or two.

Homework for Week 2

Problem 1.

Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

Problem 2.



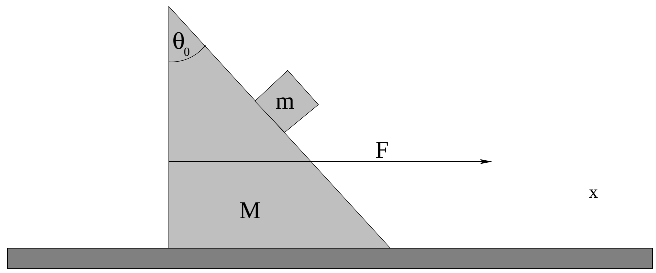
Two blocks, each with the same mass m but made of different materials, sit on a rough plane inclined at an angle θ such that they will slide (so that the component of their weight down the incline exceeds the maximum force exerted by static friction). The first (upper) block has a coefficient of *kinetic* friction of μ_{1k} between block and inclined plane; the second (lower) block has coefficient of kinetic friction μ_{2k}. The two blocks are connected by an Acme (massless and unstretchable) string. Ignore atmospheric drag.

Find the acceleration(s) of the two blocks a_1 and a_2 (which might be the same or might be different!) down the incline:

a) when $\mu_{1k} > \mu_{2k}$;

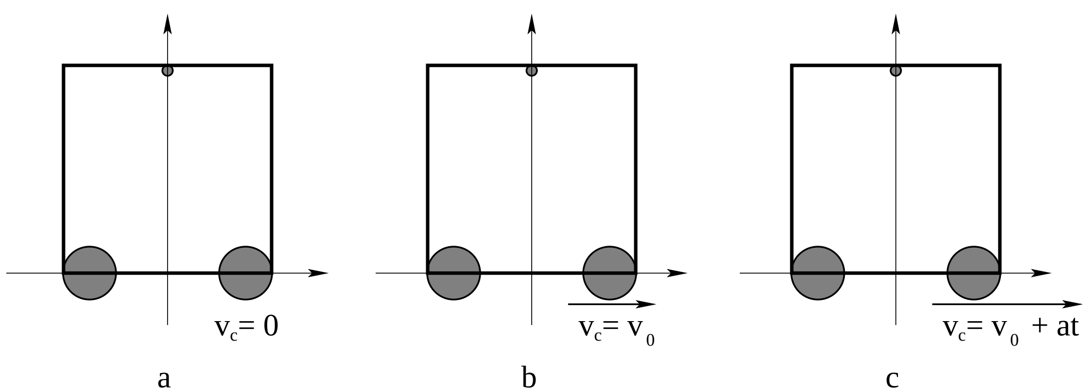
b) when $\mu_{2k} > \mu_{1k}$.

Problem 3.



A small square block m is sitting on a larger wedge-shaped block of mass M at an *upper* angle θ_0 such that the little block will slide on the big block if both are started from rest and no other forces are present. The large block is sitting on a frictionless table. The coefficient of static friction between the large and small blocks is μ_s. With what *range* of force F can you push on the large block to the right such that the small block will remain motionless with respect to the large block and neither slide up nor slide down?

Problem 4.



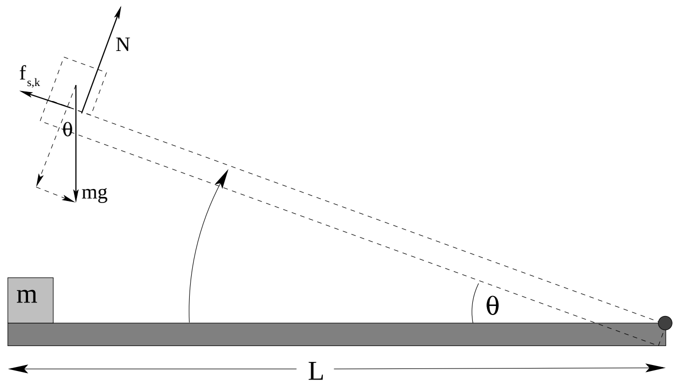
In the figure above a cart has a rubber ball suspended from its roof. At $t = 0$, the ball is ***released from rest relative to the cart*** and then falls freely under the influence of gravity. In figure a), the carts is at rest ($v_c = 0$) in the ground frame. In figure b), it is moving with a uniform speed $v_c = v_0$ to the right relative to the ground. In figure c), it starts at this same speed v_0, but accelerates to the right with acceleration a so its speed is given by $v_c = v_0 + at$ relative to the ground (where $a \approx 0.1g$).

a) Draw the expected trajectory of the ball ***in the frame of the ground*** (the provided axes) in all three cases. In particular, indicate where you think the ball is likely to hit the ground, and whether it follows a straight or curved path to get there.

b) Draw the expected trajectory of the ball ***in the frame of the cart*** (again using the provided axes) in all three cases. In particular, indicate where you think the ball is likely to hit the bottom of the cart, and whether it follows a straight or curved path to get there.

Note that no algebra is required to solve this problem. I'd suggest using different colors for the ground-frame and cart-frame trajectories, and/or clearly labeling which is which on copies of the three figures.

Problem 5.



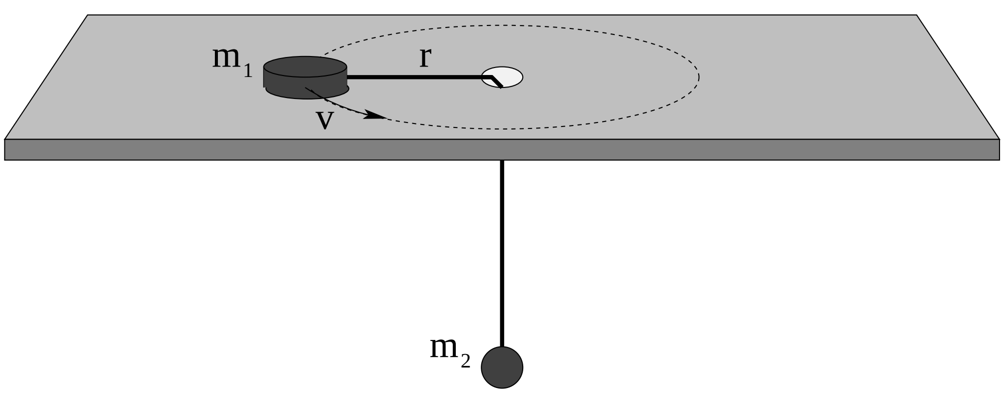
A block of mass m sits **at rest** on a rough plank of length L that can be gradually tipped up until the block slides. The coefficient of static friction between the block and the plank is μ_s; the coefficient of dynamic friction is μ_k and as usual, $\mu_k < \mu_s$.

a) Find the angle θ_0 at which the block first starts to move.

b) Suppose that the plank is lifted to an angle $\theta > \theta_0$ (where the mass will definitely slide) and the mass is released from rest at time $t_0 = 0$. Find its acceleration a down the incline.

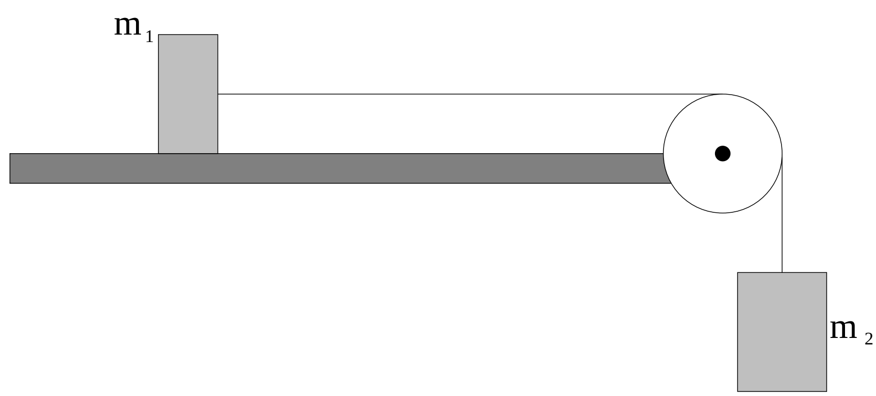
c) Finally, find the time t_f that the mass reaches the lower end of the plank.

Problem 6.



A hockey puck of mass m_1 is tied to a string that passes through a hole in a frictionless table, where it is also attached to a mass m_2 that hangs underneath. The mass is given a push so that it moves in a circle of radius r at constant speed v when mass m_2 hangs free beneath the table. Find r as a function of m_1, m_2, v, and g.

Problem 7.

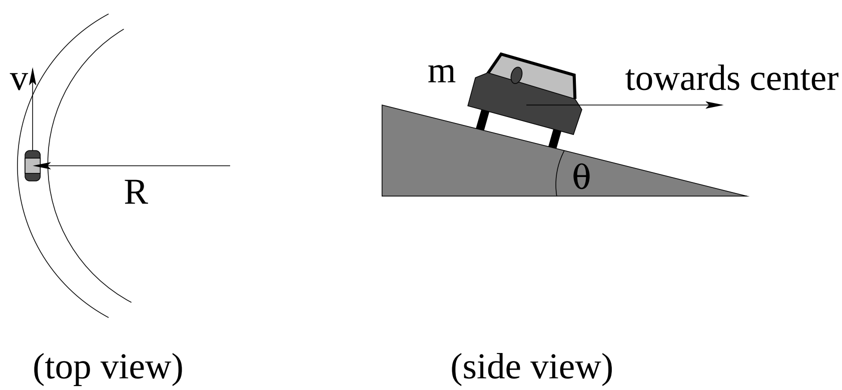


A mass m_1 is attached to a second mass m_2 by an Acme (massless, unstretchable) string. m_1 sits on a table with which it has coefficients of static and dynamic friction μ_s and μ_k respectively. m_2 is hanging over the ends of a table, suspended by the taut string from an Acme (frictionless, massless) pulley. At time $t = 0$ both masses are released.

a) What is the minimum mass $m_{2,\mathrm{min}}$ such that the two masses begin to move?

b) If $m_2 = 2m_{2,\mathrm{min}}$, determine how fast the two blocks are moving when mass m_2 has fallen a height H (assuming that m_1 hasn't yet hit the pulley)?

Problem 8.

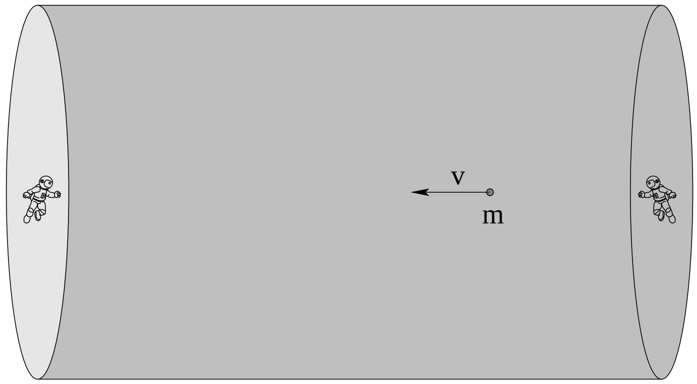


A car of mass m is rounding a banked curve that has radius of curvature R and banking angle θ. The coefficient of static friction between the car's tires and the road is μ_s.

a) Find the *range* of speeds v of the car such that it can succeed in making it around the curve without skidding.

b) Find the speed where the car successfully rounds the curve even if it is icy ("frictionless").

Problem 9.

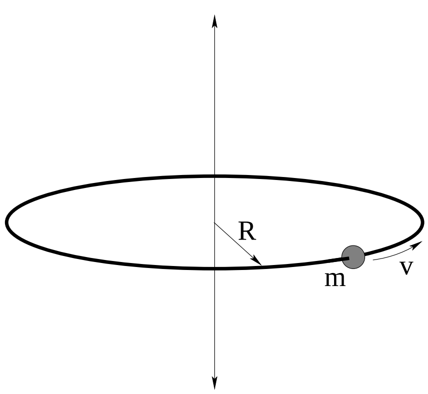


You and a friend are working inside a cylindrical new space station that is a hundred meters long and thirty meters in radius and filled with a thick air mixture. It is lunchtime and you have a bag of oranges. Your friend (working at the other end of the cylinder) wants one, so you throw one at him at speed v_0 at $t = 0$. Assume *Stokes* drag, that is $\vec{F}_d = -b\vec{v}$ (this is probably a poor assumption depending on the initial speed, but it makes the algebra relatively easy and qualitatively describes the motion well enough).

a) Derive an algebraic expression for the velocity of the orange as a function of time.

b) How long does it take the orange to lose half of its initial velocity?

Problem 10.



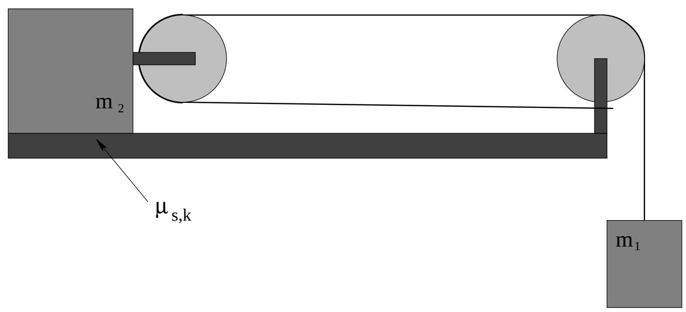
A bead of mass m is threaded on a metal hoop of radius R. There is a coefficient of kinetic friction μ_k between the bead and the hoop. It is given a push to start it sliding around the hoop with initial speed v_0. The hoop is located on the space station, so you can ignore gravity. All answers below should be given in terms of m, μ_k, R, v (where requested) and v_0.

a) Find the normal force exerted by the hoop on the bead as a function of its speed.

b) Find the dynamical frictional force exerted by the hoop on the bead as a function of its speed.

c) Find its speed as a function of time. This involves using the frictional force on the bead in Newton's second law, finding its *tangential* acceleration on the hoop (which is the time rate of change of its speed) and solving the equation of motion.

Problem 11.



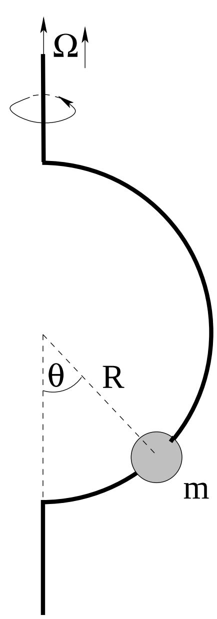
A block of mass m_2 sits on a rough table. The coefficients of friction between the block and the table are μ_s and μ_k for static and kinetic friction respectively. A mass m_1 is suspended from an massless, unstretchable, unbreakable rope that is looped around the two pulleys as shown and attached to the support of the rightmost pulley. At time $t = 0$ the system is released at rest.

a) Find an expression for the *minimum* mass $m_{1,\mathrm{min}}$ such that the masses will begin to move.

b) Suppose $m_1 = 2m_{1,\mathrm{min}}$ (twice as large as necessary to start it moving). Solve for the accelerations of *both masses.* Hint: Is there a constraint between how far mass m_2 moves when mass m_1 moves down a short distance?

c) Find the speed of both masses after the small mass has fallen a distance H. Remember this answer and how hard you had to work to find it – next week we will find it much more easily.

Advanced Problem 12.



A small frictionless bead is threaded on a semicircular wire hoop with radius R, which is then spun on its vertical axis as shown above at angular velocity $\vec{\Omega}$.

a) Find an expression for θ in terms of R, g and (angular speed) $\Omega = |\vec{\Omega}|$.

b) What is the *smallest angular speed* Ω_{min} such that the bead will not sit at the bottom at $\theta = 0$, for a given R.

Week 3: Work and Energy

1.9: Summary

- **The Work-Kinetic Energy Theorem** in words is "The work done by the total force acting on an object between two points equals the change in its kinetic energy." As is frequently the case, though, this is more usefully learned in terms of its algebraic forms:

$$W(x_1 \to x_2) = \int_{x_1}^{x_2} F_x dx = \frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2 = \Delta K \tag{3.1}$$

  in one dimension or

$$W(\vec{\boldsymbol{x}}_1 \to \vec{\boldsymbol{x}}_2) = \int_{x_1}^{x_2} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{\ell}} = \frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2 = \Delta K \tag{3.2}$$

  in two or more dimensions, where the integral in the latter is ***along some specific path*** between the two endpoints.

- **A Conservative Force** $\vec{\boldsymbol{F}}_c$ is one where the integral:

$$W(\vec{\boldsymbol{x}}_1 \to \vec{\boldsymbol{x}}_2) = \int_{x_1}^{x_2} \vec{\boldsymbol{F}}_c \cdot d\vec{\boldsymbol{\ell}} \tag{3.3}$$

  does not depend on the particular path taken between $\vec{\boldsymbol{x}}_1$ and $\vec{\boldsymbol{x}}_2$. In that case going from $\vec{\boldsymbol{x}}_1$ to $\vec{\boldsymbol{x}}_2$ by one path and coming back by another forms a *loop* (a closed curve containing both points). We must do the same amount of positive work going one way as we do negative the other way and therefore we can write the condition as:

$$\oint_C \vec{\boldsymbol{F}}_c \cdot d\vec{\boldsymbol{\ell}} = 0 \tag{3.4}$$

  for all closed curves C.

  Note Well: If you have no idea what the dot-product in these equations is or how to evaluate it, if you don't know what an integral along a curve is, it might be a good time to go over the former in the online math review and pay close attention to the pictures below that explain it in context. Don't *worry* about this – it's all part of what you need to learn in the course, and I don't expect that you have a particularly good grasp of it *yet*, but it is definitely something to work on!

- **Potential Energy** is the negative work done by a **conservative force** (only) moving between two points. The reason that we bother defining it is because for known, conservative force rules, we can do the work integral *once and for all* for the functional form of the force and obtain an answer that is (within a constant) the *same for all problems!* We can then simplify the Work-Kinetic Energy Theorem for problems involving those conservative forces, changing them into *energy conservation problems* (see below). Algebraically:

$$U(\vec{\boldsymbol{x}}) = -\int \vec{\boldsymbol{F}}_c \cdot d\vec{\boldsymbol{\ell}} + U_0 \tag{3.5}$$

where the integral is the *indefinite* integral of the force and U_0 is an arbitrary constant of integration (that may be set by some convention though it doesn't really have to be, be wary) or else the change in the potential energy is:

$$\Delta U(\vec{\boldsymbol{x}}_0 \to \vec{\boldsymbol{x}}_1) = -\int_{\vec{\boldsymbol{x}}_0}^{\vec{\boldsymbol{x}}_1} \vec{\boldsymbol{F}}_c \cdot d\vec{\boldsymbol{\ell}} \tag{3.6}$$

(independent of the choice of path between the points).

- The **Law of Conservation of Mechanical Energy** states that if no non-conservative forces are acting, the sum of the potential and kinetic energies of an object are *constant* as the object moves around:

$$E_i = U_0 + K_0 = U_f + K_f = E_f \tag{3.7}$$

  where $U_0 = U(\vec{\boldsymbol{x}}_0)$, $K_0 = \frac{1}{2}mv_0^2$ etc.

- The **Generalized Non-Conservative Work-Mechanical Energy Theorem** states that if both conservative and non-conservative forces are acting on an object (particle), the work done by the non-conservative forces (e.g. friction, drag) equals the change in the total mechanical energy:

$$W_{nc} = \int_{\vec{\boldsymbol{x}}_0}^{\vec{\boldsymbol{x}}_1} \vec{\boldsymbol{F}}_{nc} \cdot d\vec{\boldsymbol{\ell}} = \Delta E_{\rm mech} = (U_f + K_f) - (U_0 + K_0) \tag{3.8}$$

  In general, recall, the work done by non-conservative forces **depends on the path taken**, so the left hand side of this must be explicitly evaluated for a particular path while the right hand side depends only on the values of the functions at the end points of that path.

  Note well: This is a theorem only if one considers the ***external*** forces acting on a ***particle***. When one considers systems of particles or objects with many "internal" degrees of freedom, things are not this simple because there can be non-conservative internal forces that (for example) can add or remove ***macroscopic*** mechanical energy to/from the system and turn it into ***microscopic*** mechanical energy, for example chemical energy or "heat". Correctly treating energy at this level of detail requires us to formulate ***thermodynamics*** and is beyond the scope of the current course, although it requires a good understanding of its concepts to get started.

- **Power** is the work performed per unit time by a force:

$$P = \frac{dW}{dt} \tag{3.9}$$

  In many mechanics problems, power is most easily evaluated by means of:

$$P = \frac{d}{dt}\left(\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{\ell}}\right) = \vec{\boldsymbol{F}} \cdot \frac{d\vec{\boldsymbol{\ell}}}{dt} = \vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{v}} \tag{3.10}$$

- An object is in *force equilibrium* when its potential energy function is at a **minimum** or **maximum**. This is because the other way to write the definition of potential energy is:
$$F_x = -\frac{dU}{dx} \tag{3.11}$$
so that if
$$\frac{dU}{dx} = 0 \tag{3.12}$$
then $F_x = 0$, the condition for force equilibrium in one dimension.

  For advanced students: In more than one dimension, the force is the negative *gradient* of the potential energy:
$$\vec{\boldsymbol{F}} = -\vec{\boldsymbol{\nabla}}U = -\frac{\partial U}{\partial x}\hat{\boldsymbol{x}} - \frac{\partial U}{\partial y}\hat{\boldsymbol{y}} - \frac{\partial U}{\partial z}\hat{\boldsymbol{z}} \tag{3.13}$$
(where $\frac{\partial}{\partial x}$ stands for the *partial derivative* with respect to x, the derivative of the function one takes pretending the other coordinates are constant.

- An equilibrium point $\vec{\boldsymbol{x}}_e$ is **stable** if $U(\vec{\boldsymbol{x}}_e)$ is a *minimum*. A mass hanging at rest from a string is at a stable equilibrium at the bottom.

- An equilibrium point $\vec{\boldsymbol{x}}_e$ is **unstable** if $U(\vec{\boldsymbol{x}}_e)$ is a *maximum*. A pencil balanced on its point (if you can ever manage such a feat) is in unstable equilibrium – the slightest disturbance and it will fall.

- An equilibrium point $\vec{\boldsymbol{x}}_e$ is **neutral** if $U(\vec{\boldsymbol{x}}_e)$ is flat to either side, neither ascending or descending. A disk placed on a perfectly level frictionless table is in neutral equilibrium – *if* it is place at rest, it will remain at rest no matter where you place it, but of course if it has the slightest nonzero velocity it will coast until it either reaches the edge of the table or some barrier that traps it. In the latter sense a perfect neutral equilibrium is often really unstable, as it is essentially impossible to place an object at rest, but friction or drag often conspire to "stabilize" a neutral equilibrium so that yes, if you put a penny on a table it will be there the next day, unmoved, as far as *physics* is concerned...

3.1: Work and Kinetic Energy

If you've been doing all of the work assigned so far, you may have noticed something. In *many* of the problems, you were asked to find the *speed* of an object (or, if the direction was obvious, its velocity) after it moved from some initial position to a final position. The solution strategy you employed over and over again was to solve the equations of motion, solve for the time, substitute the time, find the speed or velocity. We used this in the very first example in the book and the first actual homework problem to show that a mass dropped from rest that falls a height H hits the ground at speed $v = \sqrt{2gH}$, but later we discovered that a mass that slides down a frictionless inclined plane starting from rest a height H above the ground arrives at the ground as a speed $\sqrt{2gH}$ *independent of the slope of the incline!*

If you were mathematically inclined – or used a different textbook, one with a separate section on the kinematics of constant acceleration motion (a subject this textbook has assiduously avoided, instead requiring you to actually *solve the equations of motion using calculus* repeatedly and then use algebra as needed to answer the questions) you might have noted that you can actually do the algebra associated with this elimination of time *once and for all* for a constant acceleration problem in one dimension. It is simple.

If you look back at week 1, you can see if that if you integrate a constant acceleration of an object twice, you obtain:

$$\begin{aligned} v(t) &= at + v_0 \\ x(t) &= \frac{1}{2}at^2 + v_0 t + x_0 \end{aligned}$$

as a completely general kinematic solution in one dimension, where v_0 is the initial speed and x_0 is the initial x position at time $t = 0$.

Now, suppose you want to find the speed v_1 the object will have when it reaches position x_1. One can *algebraically, once and for all* note that this must occur at some time t_1 such that:

$$\begin{aligned} v(t_1) &= at_1 + v_0 = v_1 \\ x(t_1) &= \frac{1}{2}at_1^2 + v_0 t_1 + x_0 = x_1 \end{aligned}$$

We can algebraically solve the first equation *once and for all* for t_1:

$$t_1 = \frac{v_1 - v_0}{a} \tag{3.14}$$

and substitute the result into the second equation, elminating time altogether from the solutions:

$$\begin{aligned} \frac{1}{2}a\left(\frac{v_1 - v_0}{a}\right)^2 + v_0\left(\frac{v_1 - v_0}{a}\right) + x_0 &= x_1 \\ \frac{1}{2a}\left(v_1^2 - 2v_0v_1 + v_0^2\right) + \left(\frac{v_0v_1 - v_0^2}{a}\right) &= x_1 - x_0 \\ v_1^2 - 2v_0v_1 + v_0^2 + 2v_0v_1 - 2v_0^2 &= 2a(x_1 - x_0) \end{aligned}$$

or

$$v_1^2 - v_0^2 = 2a(x_1 - x_0) \tag{3.15}$$

Many textbooks encourage students to memorize this equation as well as the two kinematic solutions for constant acceleration very early – often before one has even learned Newton's Laws – so that students never have to actually learn *why* these solutions are important or *where they come from*, but at this point you've hopefully learned both of those things well and it is time to make solving problems of this kinds a little bit easier.

However, we will not do so using this constant acceleration kinematic equation even now! There is no need! As we will see below, it is quite simple to eliminate time from *Newton's Second Law itself* once and for all, and obtain a powerful way of solving many, many physics problems – in particular, ones where the questions asked do not depend on specific times – *without* the tedium of integrating out the equations of motion. This "time independent" formulation of force laws and motion turns out, in the end, to be even more general and useful than Newton's Laws themselves, surviving the transition to quantum theory where the concepts of force and acceleration do not.

One *very good thing* about waiting as we have done and not memorizing anything, let alone kinematic constant acceleration solutions, is that this new formulation in terms of *work* and *energy* works just fine for *non-constant* forces and accelerations, where the kinematic solutions above are (as by now you should fully appreciate, having worked through e.g. the drag force and investigated the force exerted by springs, neither of which are constant in space or in time) completely useless and wrong.

Let us therefore begin *now* with this relatively meaningless kinematical result that arises from eliminating time for a ***constant acceleration in one dimension only*** – planning to use it only long enough to ensure that we never have to use it because we've found something even better that is far more *meaningful*:

$$v_1^2 - v_0^2 = 2a\Delta x \tag{3.16}$$

where Δx is the displacement of the object $x_1 - x_0$.

If we multiply by m (the mass of the object) and move the annoying 2 over to the other side, we can make the constant acceleration a into a constant force $F_x = ma$:

$$\begin{aligned} (ma)\Delta x &= \frac{1}{2}mv_1^2 - \frac{1}{2}mv_0^2 & (3.17)\\ F_x\Delta x &= \frac{1}{2}mv_1^2 - \frac{1}{2}mv_0^2 & (3.18) \end{aligned}$$

We now *define* the ***work done by the constant force*** F_x ***on the mass*** m as it moves through the distance Δx to be:

$$\Delta W = F_x\Delta x. \tag{3.19}$$

The work can be positive or negative.

Of course, **not all forces are constant**. We have to wonder, then, if this result or concept is as fragile as the integral of a constant acceleration (which does not "work", so to speak, for springs!) or if it can handle springs, pendulums, real gravity (not near the Earth's surface) and so on. As you might guess, the answer is yes – we wouldn't have bothered introducing and naming the concept if all we cared about was constant acceleration problems as we already had a satisfactory solution for them – but before we turn this initial result into a *theorem* that follows directly from the axiom of *Newton's Second Law made independent of time*, we should discuss units of work, energy, and all that.

3.1.1: Units of Work and Energy

Work is a form of ***energy***. As always when we first use a new named quantity in physics, we need to define its *units* so we can e.g. check algebraic results for kinematic consistency, correctly identify work, and learn to quantitatively appreciated it when people refer to quantities in other sciences or circumstances (such as the energy yield of a chemical reaction, the power consumed by an electric light bulb, or the energy consumed and utilized by the human body in a day) in these units.

In general, the definition of SI units can most easily be remembered and understood from the basic equations that define the quantity of interest, and the units of energy are no exception. Since work is defined above to be a force times a distance, the SI units of energy must be the SI units of force (Newtons) times the SI units of length (meters). The units themselves are named (as many are) after a Famous Physicist, James Prescott Joule[97] . Thus:

$$1 \text{ Joule} = 1 \text{ Newton-meter} = 1 \, \frac{\text{kilogram-meter}^2}{\text{second}^2} \tag{3.20}$$

3.1.2: Kinetic Energy

The latter, we also note, are the natural units of mass times speed squared. We observe that this is the quantity that *changes* when we do work on a mass, and that this energy appears to be a characteristic of the moving mass associated with the motion itself (dependent only on the speed v). We therefore define the quantity changed by the work to be the ***kinetic energy***[98] and will use the symbol K to represent it in this work:

$$K = \frac{1}{2} m v^2 \tag{3.21}$$

Note that kinetic energy is a **relative** quantity – it depends upon the inertial frame in which it is measured. Suppose we consider the kinetic energy of a block of mass m sliding forward at a constant speed v_b in a railroad car travelling at a constant speed v_c. The frame of the car is an *inertial reference frame* and so Newton's Laws must be valid there. In particular, our definition of kinetic energy that followed from a solution to Newton's Laws ought to be valid there. It should be equally valid on the ground, but these two quantities are *not equal*.

Relative to the ground, the speed of the block is:

$$v_g = v_b + v_c \tag{3.22}$$

and the kinetic energy of the block is:

$$K_g = \frac{1}{2} m v_g^2 = \frac{1}{2} m v_b^2 + \frac{1}{2} m v_c^2 + m v_b v_c \tag{3.23}$$

or

$$K_g = K_b + \frac{1}{2} m v_c^2 + m v_b v_c \tag{3.24}$$

where K_b is the kinetic energy of the block in the frame of the train.

Worse, the train is riding on the Earth, which is not exactly at rest relative to the sun, so we could describe the velocity of the block by adding the velocity of the Earth to that of the train and the block within the train. The kinetic energy in this case is so large that the *difference* in the energy of the block due to its relative motion in the train coordinates is almost invisible against the huge energy it has in an inertial frame in which the sun is approximately at rest. Finally, as we discussed last week, the sun itself is moving relative to the galactic center or the "rest frame of the visible Universe".

---

[97] Wikipedia: http://www.wikipedia.org/wiki/James Prescott Joule. He worked with temperature and heat and was one of the first humans on Earth to formulate and initially experimentally verify the Law of Conservation of Energy, discussed below. He also discovered and quantified resistive electrical heating (Joule heating) and did highly precise experiments that showed that mechanical energy delivered into a closed system increased its temperature is the work converted into heat.

[98] The work "kinetic" means "related to the motion of material bodies", although we also apply it to e.g. hyperkinetic people...

What, then, is the *actual* kinetic energy of the block?

I don't know that there is such a thing. But the kinetic energy of the block in the inertial reference frame of *any well-posed problem* is $\frac{1}{2}mv^2$, and that will have to be enough for us. As we will prove below, this definition makes the work done by the forces of nature *consistent* within the frame, so that our computations will give us answers consistent with experiment and experience in the frame coordinates.

3.2: The Work-Kinetic Energy Theorem

Let us now formally state the result we derived above using the new definitions of work and kinetic energy as the ***Work-Kinetic Energy Theorem*** (which I will often abbreviate, e.g. **WKET**) in one dimension in English:

> The work done on a mass by the total force acting on it is equal to the change in its kinetic energy.

and as an equation that is correct for constant one dimensional forces only:

$$\Delta W = F_x \Delta x = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2 = \Delta K \tag{3.25}$$

You will note that in the English statement of the theorem, I didn't say anything about needing the force to be constant or one dimensional. I did this because those conditions *aren't necessary* – I only used them to bootstrap and motivate a *completely general result*. Of course, now it is up to us to *prove* that the theorem is general and determine its correct and general algebraic form. We can, of course, guess that it is going to be the integral of this *difference* expression turned into a *differential* expression:

$$dW = F_x dx = dK \tag{3.26}$$

but actually deriving this for an explicitly non-constant force has several important conceptual lessons buried in the derivation. So much so that I'll derive it in *two completely different ways.*

3.2.1: Derivation I: Rectangle Approximation Summation

First, let us consider a force that varies with position so that it can be mathematically described as a function of x, $F_x(x)$. To compute the work done going between (say) x_0 and some position x_f that will ultimately equal the total change in the kinetic energy, we can try to chop the interval $x_f - x_0$ up into lots of small pieces, each of width Δx. Δx needs to be small enough that F_x basically doesn't change much across it, so that we are justified in saying that it is "constant" across each interval, even though the value of the constant isn't exactly the same from interval to interval. The actual value we use as the constant in the interval isn't terribly important – the easiest to use is the average value or value at the midpoint of the interval, but no matter what sensible value we use the error we make will vanish as we make Δx smaller and smaller.

In figure 3.1, you can see a very crude sketch of what one might get chopping the total interval $x_0 \to x_f$ up into eight pieces such that e.g. $x_1 = x_0 + \Delta x$, $x_2 = x_1 + \Delta x$,... and

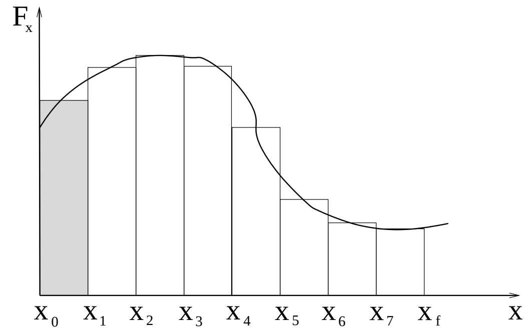


Figure 3.1: The work done by a variable force can be approximated arbitrarily accurately by summing $F_x \Delta x$ using the *average* force as if it were a constant force across each of the "slices" of width Δx one can divide the entire interval into. In the limit that the width $\Delta x \to dx$, this summation turns into the *integral*.

computing the work done across *each* sub-interval using the approximately constant value it has in the middle of the sub-interval. If we let $F_1 = F_x(x_0 + \Delta x/2)$, then the work done in the first interval, for example, is $F_1 \Delta x$, the shaded area in the first rectangle draw across the curve. Similarly we can find the work done for the second strip, where $F_2 = F_x(x_1 + \Delta x/2)$ and so on. In each case the work done equals the change in kinetic energy as the particle moves across each interval from x_0 to x_f.

We then *sum* the constant acceleration Work-Kinetic-Energy theorem for all of these intervals:

$$\begin{aligned} W_{\rm tot} &= F_1(x_1 - x_0) + F_2(x_2 - x_1) + \ldots \\ &= (\frac{1}{2}mv^2(x_1) - \frac{1}{2}mv^2(x_0)) + (\frac{1}{2}mv^2(x_2) - \frac{1}{2}mv^2(x_1)) + \ldots \\ F_1\Delta x + F_2 \Delta x + \ldots &= \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2 \\ \sum_{i=1}^{8} F_i \Delta x &= \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2 \end{aligned} \tag{3.27}$$

where the internal changes in kinetic energy at the end of each interval but the first and last **cancel**. Finally, we let Δx go to zero in the usual way, and replace summation by integration. Thus:

$$W_{\rm tot} = \lim_{\Delta x \to 0} \sum_{i=1}^{\infty} F_x(x_0 + i\Delta x)\Delta x = \int_{x_0}^{x_f} F_x dx = \Delta K \tag{3.28}$$

and we have generalized the theorem to include non-constant forces in one dimension[99].

This approach is good in that it makes it very clear that the work done is the area under the curve $F_x(x)$, but it buries the key idea – the elimination of time in Newton's Second Law – way back in the derivation and relies uncomfortably on constant force/acceleration results. It is much more elegant to directly derive this result using straight up calculus, and honestly it is a lot easier, too.

[99]This is notationally a bit sloppy, as I'm not making it clear that as Δx gets smaller, you have to systematically increase the number of pieces you divide $x_f - x_0$ into and so on, hoping that you all remember your intro calculus course and recognize this picture as being one of the first things you learned about integration...

3.2.2: Derivation II: Calculus-y (Chain Rule) Derivation

To do that, we simply take Newton's Second Law and eliminate dt using the chain rule. The algebra is:

$$\begin{aligned}
F_x &= ma_x = m\frac{dv_x}{dt} \\
F_x &= m\frac{dv_x}{dx}\frac{dx}{dt} \qquad \text{(chain rule)} \\
F_x &= m\frac{dv_x}{dx}v_x \qquad \text{(definition of } v_x\text{)} \\
F_x dx &= mv_x dv_x \qquad \text{(rearrange)} \\
\int_{x_0}^{x_1} F_x dx &= m\int_{v_0}^{v_1} v_x dv_x \qquad \text{(integrate both sides)} \\
W_{\text{tot}} = \int_{x_0}^{x_1} F_x dx &= \frac{1}{2}mv_1^2 - \frac{1}{2}mv_0^2 \qquad \text{(The WKE Theorem)}
\end{aligned} \tag{3.29}$$

This is an *elegant* proof, one that makes it completely clear that time dependence is being eliminated in favor of the direct dependence of v on x. It is also clearly valid for very general one dimensional force functions; at no time did we assume anything about F_x other than its general integrability in the last step.

What happens if $\vec{F}$ is actually a *vector* force, not necessarily in acting only in one dimension? Well, the first proof above is clearly valid for $F_x(x)$, $F_y(y)$ and $F_z(z)$ independently, so:

$$\int \vec{F} \cdot d\vec{\ell} = \int F_x dx + \int F_y dy + \int F_z dz = \Delta K_x + \Delta K_y + \Delta K_z = \Delta K \tag{3.30}$$

However, this doesn't make the *meaning* of the integral on the left very clear.

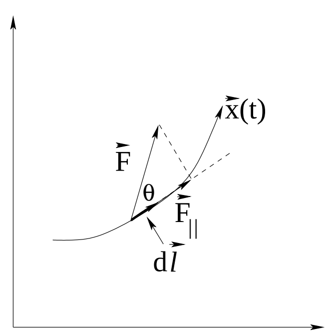


Figure 3.2: Consider the work done going along the particular trajectory $\vec{x}(t)$ where there is a force $\vec{F}(\vec{x})$ acting on it that varies along the way. As the particle moves across the small/differential section $d\vec{\ell}$, only the force component *along* $d\vec{\ell}$ does work. The other force component changes the *direction* of the velocity without changing its *magnitude*.

The best way to understand that is to examine a small piece of the path in two dimensions. In figure 3.2 a small part of the trajectory of a particle is drawn. A small chunk of that trajectory $d\vec{\ell}$ represents the vector displacement of the object over a very short time under the action of the force $\vec{F}$ acting there.

The component of $\vec{F}$ perpendicular to $d\vec{\ell}$ doesn't change the *speed* of the particle; it behaves like a centripetal force and alters the direction of the velocity without altering the speed.

The component *parallel* to $d\vec{\boldsymbol{\ell}}$, however, *does* alter the speed, that is, does work. The magnitude of the component in this direction is (from the picture) $F\cos(\theta)$ where θ is the angle between the direction of $\vec{\boldsymbol{F}}$ and the direction of $d\vec{\boldsymbol{\ell}}$.

That component acts (over this very short distance) like a one dimensional force in the direction of motion, so that

$$dW = F\cos(\theta)d\ell = d(\frac{1}{2}mv^2) = dK \tag{3.31}$$

The next little chunk of $\vec{\boldsymbol{x}}(t)$ has a different force and direction but the form of the work done and change in kinetic energy as the particle moves over that chunk is the same. As before, we can integrate from one end of the path to the other doing only the *one* dimensional integral of the path element $d\ell$ times $F_{||}$, the component of $\vec{\boldsymbol{F}}$ parallel to the path at that (each) point.

The vector operation that multiplies a vector by the component of another vector in the same direction as the first is the ***dot (or scalar) product***. The dot product between two vectors $\vec{\boldsymbol{A}}$ and $\vec{\boldsymbol{B}}$ can be written more than one way (all equally valid):

$$\begin{aligned} \vec{\boldsymbol{A}}\cdot\vec{\boldsymbol{B}} &= AB\cos(\theta) &(3.32)\\ &= A_xB_x + A_yB_y + A_zB_z &(3.33)\end{aligned}$$

The second form is connected to what we got above just adding up the independent cartesian component statements of the Work-Kinetic Energy Theorem in one (each) dimension, but it doesn't help us understand how to do the integral between specific starting and ending coordinates along some trajectory. The first form of the dot product, however, corresponds to our picture:

$$dW = F\cos(\theta)d\ell = \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{\ell}} = dK \tag{3.34}$$

Now we can see what the integral means. We have to sum this up *along some specific path between $\vec{\boldsymbol{x}}_0$ and $\vec{\boldsymbol{x}}_1$* to find the total work done moving the particle along that path by the force. For differential sized chunks, the "sum" becomes an integral and we integrate this along the path to get the correct statement of the Work-Kinetic Energy Theorem in 2 or 3 dimensions:

$$W(\vec{\boldsymbol{x}}_0 \to \vec{\boldsymbol{x}}_1) = \int_{\vec{\boldsymbol{x}}_0,\mathrm{path}}^{\vec{\boldsymbol{x}}_1} \vec{\boldsymbol{F}}\cdot d\vec{\boldsymbol{\ell}} = \frac{1}{2}mv_1^2 - \frac{1}{2}mv_0^2 = \Delta K \tag{3.35}$$

Note well that this integral may well be *different for different paths* connecting points $\vec{\boldsymbol{x}}_0$ to $\vec{\boldsymbol{x}}_1$! In the most general case, one cannot find the work done without knowing the path taken, because there are many ways to go between any two points and the forces could be very different along them.

Note well: Energy is a **scalar** – just a number with a magnitude and units but no *direction* – and hence is considerably easier to treat than vector quantities like forces.

Note well: Normal forces (perpendicular to the direction of motion) ***do no work***:

$$\Delta W = \vec{\boldsymbol{F}}_\perp \cdot \Delta\vec{\boldsymbol{x}} = 0. \tag{3.36}$$

In fact, force components perpendicular to the trajectory *bend* the trajectory with local curvature $F_\perp = mv^2/R$ but don't speed the particle up or slow it down. This *really* simplifies problem solving, as we shall see.

We should think about using time-independent work and energy instead of time dependent Newtonian dynamics whenever the answer requested for a given problem is *independent of time*. The reason for this should now be clear: we *derived* the work-energy theorem (and energy conservation) from the elimination of t from the dynamical equations.

Let's look at a few examples to see how work and energy can make our problem solving lives *much* better.

Example 3.2.1: Pulling a Block

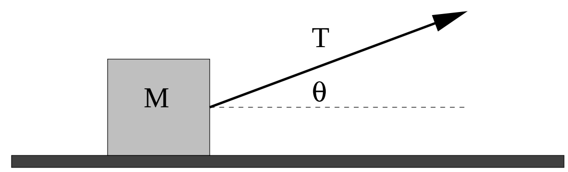


Figure 3.3: A block is connected to an Acme (massless, unstretchable) string and is pulled so that it exerts a constant tension T on the block at the angle θ.

Suppose we have a block of mass m being pulled by a string at a constant tension T at an angle θ with the horizontal along a rough table with coefficients of friction $\mu_s > \mu_k$. Typical questions might be: At what value of the tension does the block begin to move? If it is pulled with exactly that tension, how fast is it moving after it is pulled a horizontal distance L?

We see that in the y-direction, $N + T\sin(\theta) - mg = 0$, or $N = mg - T\sin(\theta)$. In the x-direction, $F_x = T\cos(\theta) - \mu_s N = 0$ (at the point where block barely begins to move).

Therefore:

$$T\cos(\theta) - \mu_s mg + T\mu_s \sin(\theta) = 0 \tag{3.37}$$

or

$$T = \frac{\mu_s mg}{\cos(\theta) + \mu_s \sin(\theta)} \tag{3.38}$$

With this value of the tension T, the work energy theorem becomes:

$$W = F_x L = \Delta K \tag{3.39}$$

where $F_x = T\cos(\theta) - \mu_k(mg - T\sin(\theta)$. That is:

$$\left(T\cos(\theta) - \mu_k(mg - T\sin(\theta)\right) L = \frac{1}{2}mv_f^2 - 0 \qquad \text{(since } v_i = 0) \tag{3.40}$$

or (after a bit of algebra, substituting in our value for T from the first part):

$$v_f = \left(\frac{2\mu_s gL\cos(\theta)}{\cos(\theta) + \mu_s \sin(\theta)} - 2\mu_k gL + \frac{2\mu_k\mu_s gL\sin(\theta)}{\cos(\theta) + \mu_s \sin(\theta)} \right)^{\frac{1}{2}} \tag{3.41}$$

Although it is difficult to check exactly, we can see that if $\mu_k = \mu_s$, $v_f = 0$ (or the mass doesn't accelerate). This is consistent with our value of T – the value at which the mass will exactly not move against μ_s alone, but will still move if "tapped" to get it started so that static friction falls back to weaker dynamic friction.

This is an example of how we can combine Newton's Laws or statics with work and energy for different parts of the same problem. So is the next example:

Example 3.2.2: Range of a Spring Gun

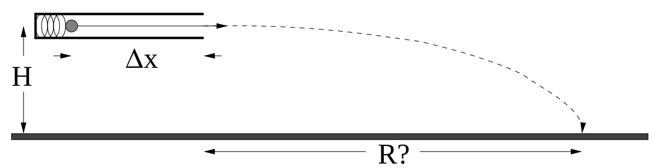


Figure 3.4: A simple spring gun is fired horizontally a height H above the ground. Compute its range R.

Suppose we have a spring gun with a bullet of mass m compressing a spring with force constant k a distance Δx. When the trigger is pulled, the bullet is released from rest. It passes down a horizontal, frictionless barrel and comes out a distance H above the ground. What is the range of the gun?

If we knew the speed that the bullet had coming out of the barrel, we'd know exactly how to solve this as in fact we *have* solved it for homework (although you shouldn't look – see if you can do this on your own or anticipate the answer below for the extra practice and review). To find that speed, we can use the Work-Kinetic Energy Theorem *if* we can compute the work done by the spring!

So our first chore then is to compute the work done by the spring that is initially compressed a distance Δx, and use that in turn to find the speed of the bullet leaving the barrel.

$$\begin{aligned} W &= \int_{x_1}^{x_0} -k(x - x_0)dx && (3.42) \\ &= -\frac{1}{2}k(x - x_0)^2|_{x_1}^{x_0} && (3.43) \\ &= \frac{1}{2}k(\Delta x)^2 = \frac{1}{2}mv_f^2 - 0 && (3.44) \end{aligned}$$

or

$$v_f = \sqrt{\frac{k}{m}}\,|\Delta x| \tag{3.45}$$

As you can see, this was pretty easy. It is also a result that we *can get no other way*, so far, because we *don't know how to solve the equations of motion for the mass on the spring* to find $x(t)$, solve for t, find $v(t)$, substitute to find v and so on. If we hadn't derived the WKE theorem for non-constant forces we'd be screwed!

The rest should be familiar. Given this speed (in the x-direction), find the range from Newton's Laws:

$$\vec{F} = -mg\hat{y} \tag{3.46}$$

or $a_x = 0$, $a_y = -g$, $v_{0x} = v_f$, $v_{0y} = 0$, $x_0 = 0$, $y_0 = H$. Solving as usual, we find:

$$R = v_{x0} t_0 \tag{3.47}$$

$$= v_f \sqrt{\frac{2H}{g}} \tag{3.48}$$

$$= \sqrt{\frac{2kH}{mg}} |\Delta x| \tag{3.49}$$

where you can either fill in the details for yourself or look back at your homework. Or get help, of course. If you can't do this second part on your own at this point, you probably *should* get help, seriously.

3.3: Review of Multiplication by a Scalar and the Dot Product

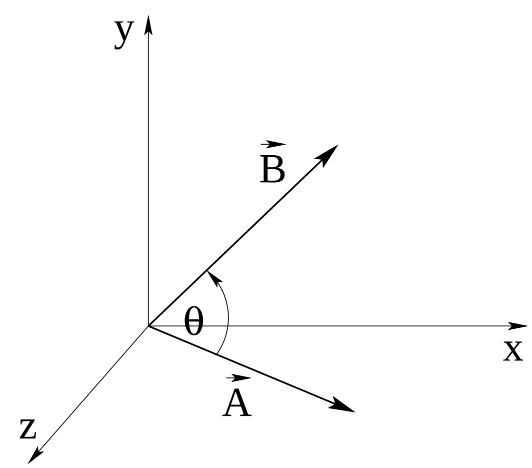


Figure 3.5: Two vectors in a three-dimensional (cartesian) space, illustrating the angle between the two vectors

Students entering the course often (very often, according to my direct measurements by means of a pre-class assessment) are unfamiliar with the principle ways of multiplying vectors – multiplication of a vector *by* a scalar, the dot/scalar product and the cross/vector product. This topical section will address only the first two of these – the third (cross product) will be reviewed when we reach rotation, torque and angular momentum, where it is extremely relevant.

Let's start with the standard cartesian representation of vectors $\vec{A}, \vec{B}, \ldots$:

$$\vec{A} = A_x \hat{x} + A_y \hat{y} + A_z \hat{z} \tag{3.50}$$

$$\vec{B} = B_x \hat{x} + B_y \hat{y} + B_z \hat{z} \tag{3.51}$$

$$\ldots \tag{3.52}$$

In figure 3.5, $\vec{A}$ and $\vec{B}$ in "arbitrary" directions in a three-dimensional space are illustrated, with an angle in between them. Note that the vectors all have a *magnitude* and a *direction* in the space, as they must.

The first kind of product we will often use is ***muliplication by a scalar*** (ordinary number). This kind of multiplication ***changes the magnitude (only) of a vector***, but ***does not change its direction***. Indeed, it *scales* the vector (which is why scalars are *called* scalars)!

The easiest way to understand this is to look at a few simple examples that we might encounter or have already encountered in physics (perhaps without appreciating it as a kind of product involving vectors). The first is one you may not have recognized, but is implicit in the cartesian representation of a vector in the first place!

We define $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$ (or you may have seen $\hat{\boldsymbol{i}}, \hat{\boldsymbol{j}}, \hat{\boldsymbol{k}}$ used instead) to be *unit vectors in the three cartesian directions*. The vector $\vec{\boldsymbol{A}}$ is the sum of *three perpendicular vectors*, each represented by *scaling* the unit vector by the magnitudes of the vector components, A_x, A_y, and A_z:

$$\vec{\boldsymbol{A}}_x = A_x\hat{\boldsymbol{x}} \tag{3.53}$$

$$\vec{\boldsymbol{A}}_y = A_y\hat{\boldsymbol{y}} \tag{3.54}$$

$$\vec{\boldsymbol{A}}_z = A_z\hat{\boldsymbol{z}} \tag{3.55}$$

Thus $\vec{\boldsymbol{A}}$ is the sum of *three* vectors:

$$\vec{\boldsymbol{A}} = \vec{\boldsymbol{A}}_x + \vec{\boldsymbol{A}}_y + \vec{\boldsymbol{A}}_z = A_x\hat{\boldsymbol{x}} + A_y\hat{\boldsymbol{y}} + A_z\hat{\boldsymbol{z}}$$

This makes it easy to see how to scale $\vec{\boldsymbol{A}}$ itself by (for example) making it twice as long:

$$\vec{\boldsymbol{A}}' = 2\vec{\boldsymbol{A}} = 2A_x\hat{\boldsymbol{x}} + 2A_y\hat{\boldsymbol{y}} + 2A_z\hat{\boldsymbol{z}} \tag{3.56}$$

We've already implicitly used this kind of product to find the vector acceleration when we divided the *vector* force by the *scalar* mass (that is *multiplied* it by the inverse of the mass) to get the *vector* acceleration! ***Note well*** that in physical kinematics (math *with units*) **scalars can be dimensioned quantities** as well as purely dimensionless numbers like 14.23 or π:

$$\vec{\boldsymbol{a}} = \frac{1}{m}\vec{\boldsymbol{F}} \tag{3.57}$$

The product we need for the WKE theorem and its variations covered in the sections ahead is the ***dot***, ***inner***, or ***scalar*** product. Don't confuse the latter with multiplying a vector by a scalar! It is named scalar product because it takes two vectors and multiplies them to *make* a scalar, not because one multiplies *by* a scalar.

In the context of the derivation, we used it to find the product of the *component* of the force $F_{||}$ in the direction of a differential step $d\vec{\boldsymbol{\ell}}$ and the magnitude of the step itself $d\ell$, *given an angle* θ in between them. This suggests that our general definition for the product of two vectors *like* this should be:

$$\vec{A} \cdot \vec{B} = A_{||}B = AB_{||} = AB\cos\theta \tag{3.58}$$

(where the components are illustrated in a plane in figure 3.6). **Note well** that I am using A (with no vector arrow over it) to stand for "the magnitude of $\vec{\boldsymbol{A}}$". This will all be consistent and end up making perfect sense below.

This is the product that we deduced the need for in the WKE.

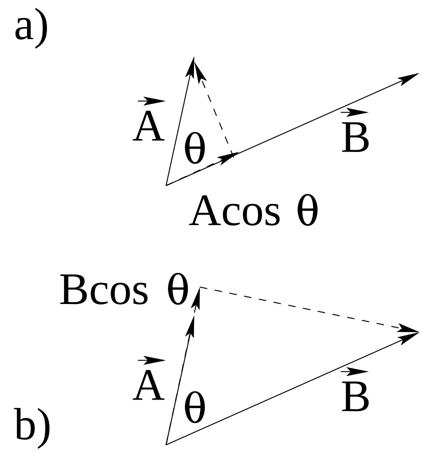


Figure 3.6: Two vectors in a plane to illustrate the first form of the dot product

This product is ***associative, distributive, and commutative***[100] It also has the following properties (definitions and/or theorems, if you like). First, the magnitude of e.g. the vector $\vec{A}$ is *defined* to be:

$$A = +\sqrt{\vec{A} \cdot \vec{A}} = +\sqrt{A^2} \tag{3.59}$$

(because $\cos 0 = 1$).

If we have two vectors $\vec{A}$ and $\vec{B}$ that are *perpendicular* (so $\theta = \pi/2$ in between them):

$$\vec{A} \cdot \vec{B} = AB \cos\frac{\pi}{2} = 0 \tag{3.60}$$

This relation is used in both directions! If the dot product of two nonzero vectors is zero, *we know that they are perpendicular!*

This is what makes the ***unit vectors*** that form what is called a ***basis*** for the cartesian frame so useful! They all have length one and an angle of $\pi/2$ between them and hence have the following multiplication table:

$$\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}} = \hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{y}} = \hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{z}} = 1 \qquad \hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = \hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{z}} = \hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{x}} = 0 \tag{3.61}$$

Note that the product is *commutative* so the last three products work the other way around as well, e.g. $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{x}} = 0$, for a total of 9 possible ways to multiply three perpendicular unit vectors in a 3 dimensional space! Cool! And *useful*...

In particular, this lets us deduce one more way of writing the dot product – we can now easily derive its form in terms of cartesian components! This lets us deduce one more way of writing the dot product in component form. Let $\vec{A} = A_x\hat{\boldsymbol{x}} + A_y\hat{\boldsymbol{y}} + A_z\hat{\boldsymbol{z}}$ and $\vec{B} = B_x\hat{\boldsymbol{x}} + B_y\hat{\boldsymbol{y}} + B_z\hat{\boldsymbol{z}}$ as usual. Then (using the ***distributivity*** of the product):

$$\begin{aligned}
\vec{A} \cdot \vec{B} = \quad & A_x B_x \hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}} + A_x B_y \cancel{\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}} + A_x B_z \cancel{\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{z}}} && (3.62)\\
& + A_y B_x \cancel{\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{x}}} + A_y B_y \hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{y}} + A_y B_z \cancel{\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{z}}} && (3.63)\\
& + A_z B_x \cancel{\hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{x}}} + A_z B_y \cancel{\hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{y}}} + A_z B_z \hat{\boldsymbol{z}} \cdot \hat{\boldsymbol{z}} && (3.64)
\end{aligned}$$

[100]Wikipedia: http://www.wikipedia.org/wiki/Multiplication#Properties. .

All of the non-diagonal dot products of unit vectors are zero (as illustrated in the cancellations above)! Hence:

$$\vec{\boldsymbol{A}} \cdot \vec{\boldsymbol{B}} = A_x B_x + A_y B_y + A_z B_z \tag{3.65}$$

and our length definition becomes:

$$A = +\sqrt{\vec{\boldsymbol{A}} \cdot \vec{\boldsymbol{A}}} = +\sqrt{A_x^2 + A_y^2 + A_z^2}$$

exactly as we might have expected from the 3D Pythagorean theorem! Indeed, this is more or less a *derivation* of the 3D Pythagorean theorem!

There are two more very useful consequences of the two products covered above that are worth mentioning before we move on. We now know how to find the magnitude of $\vec{\boldsymbol{A}}$ once we have it in a cartesian representation. But what about its direction? Can we "isolate" its direction independent of its magnitude? Yes we can! All we have to do is divide $\vec{\boldsymbol{A}}$ by its (scalar) magnitude to make a ***unit vector in the $\vec{A}$ direction***! That is:

$$\hat{\boldsymbol{A}} = \frac{\vec{\boldsymbol{A}}}{A} \qquad \Leftrightarrow \qquad \vec{\boldsymbol{A}} = A\hat{\boldsymbol{A}} \tag{3.66}$$

The second one is one of the fundamental relations underlying vector analysis and linear algebra in general. The ***correct, formal*** way of finding the component of a vector in any given direction is to ***take the dot product of the vector with a unit vector in that direction!*** I'll give two useful examples:

$$A_x = \vec{\boldsymbol{A}} \cdot \hat{\boldsymbol{x}} = \hat{\boldsymbol{x}} \cdot \vec{\boldsymbol{A}} \tag{3.67}$$

(and ditto, of course, for A_y and A_z to find the cartesian components of $\vec{\boldsymbol{A}}$). This is called ***projection***, we say that A_x is the projection of the vector $\vec{\boldsymbol{A}}$ onto the x-direction. This is because it is just like the projected "shadow" you'd get if you illuminated a vector from above to cast the shadow onto the e.g. x-axis!

Projection isn't limited to just the directions of the cartesian unit vectors. We can find the projection – the magnitude of the component of $\vec{\boldsymbol{A}}$ – onto (say) the $\vec{\boldsymbol{B}}$ direction when θ is the angle between $\vec{\boldsymbol{A}}$ and $\vec{\boldsymbol{B}}$ is illustrated above:

$$A_B = \vec{\boldsymbol{A}} \cdot \hat{\boldsymbol{B}} = \vec{\boldsymbol{A}} \cdot \frac{\vec{\boldsymbol{B}}}{B} = A\cos\theta \qquad \Leftrightarrow \qquad \vec{\boldsymbol{A}} \cdot \vec{\boldsymbol{B}} = AB\cos\theta \tag{3.68}$$

bringing us full circle around with our starting point! Everything is consistent! Note that the projection of $\vec{\boldsymbol{A}}$ onto $\vec{\boldsymbol{B}}$ is still the "shadow" of $\vec{\boldsymbol{A}}$ in the $\vec{\boldsymbol{B}}$ direction, which is, note well above, *exactly what we need for work and energy* which is why understanding all of this is *important!*

This wraps up our highly compact review of two of the three ways of multiplying vectors used extensively in this course. The third, the cross product, will be similarly reviewed in a future chapter. Hope this helps!

3.4: Conservative Forces: Potential Energy

We have now seen two kinds of forces in action. One kind is like gravity. The work done on a particle by gravity doesn't depend on the path taken through the gravitational field – it only

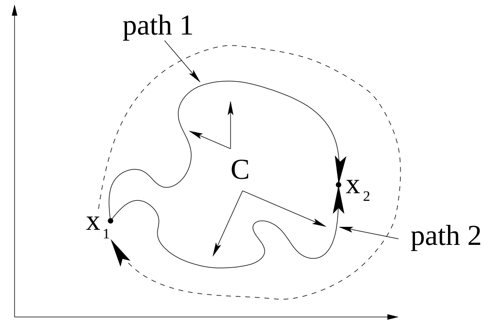


Figure 3.7: The work done going around an arbitrary loop by a conservative force is zero. This ensures that the work done going between two points is *independent* of the path taken, its defining characteristic.

depends on the relative height of the two endpoints. The other kind is like kinetic friction. Kinetic (sliding) friction not only depends on the path a particle takes, it is usually negative work; typically kinetic friction turns macroscopic mechanical energy into "heat", which can crudely be thought of an internal microscopic mechanical energy that can no longer easily be turned back into macroscopic mechanical energy. A proper discussion of heat is beyond the scope of this course, but we will remark further on this below when we discuss non-conservative forces.

We define a ***conservative force*** to be one such that the work done *by the force* as you move a point mass from point $\vec{x}_1$ to point $\vec{x}_2$ is independent of the path used to move between the points:

$$W(\vec{x}_1 \to \vec{x}_2) = \int_{\vec{x}_1 (\text{path } 1)}^{\vec{x}_2} \vec{F} \cdot d\vec{l} = \int_{\vec{x}_1 (\text{path } 2)}^{\vec{x}_2} \vec{F} \cdot d\vec{l} \tag{3.69}$$

In this case (only), the work done going around an arbitrary closed path (starting and ending on the same point) will be identically zero!

$$\begin{aligned} W_{\text{loop}} &= \oint_C \vec{F} \cdot d\vec{l} \\ &= \int_{\vec{x}_1 (\text{path } 1)}^{\vec{x}_2} \vec{F} \cdot d\vec{l} + \int_{\vec{x}_2 (\text{path } 2)}^{\vec{x}_1} \vec{F} \cdot d\vec{l} \\ &= \int_{\vec{x}_1 (\text{path } 1)}^{\vec{x}_2} \vec{F} \cdot d\vec{l} - \int_{\vec{x}_1 (\text{path } 2)}^{\vec{x}_2} \vec{F} \cdot d\vec{l} = 0 \end{aligned} \tag{3.70}$$

This is illustrated in figure 3.7. Note that the two paths from $\vec{x}_1$ to $\vec{x}_2$ combine to form a closed loop C, where the work done going forward along one path is undone coming back along the other.

We make this the *defining characteristic* of a conservative force. It is one where:

$$W_{\text{loop}} = \oint_C \vec{F}_{\text{conservative}} \cdot d\vec{l} = 0 \tag{3.71}$$

for *all closed loops one can draw in space!* This guarantees that the work done by such a force is independent of the path taken between any two points. It is *also* (in more advanced calculus) the defining characteristic of an "exact differential", the property that lets us turn it into a potential energy function below.

Since the work done moving a mass m from an arbitrary starting point to any point in space is the **same** independent of the path, we can assign each point in space a numerical value: the work done by *us* on mass m, *against* the conservative force, to reach it. This is the *negative* of the work done by the force. We do it with this sign for reasons that will become clear in a moment. We call this function the ***potential energy*** of the mass m associated with the conservative force$\vec{F}$:

$$U(\vec{x}) = -\int_{x_0}^{x} \vec{F} \cdot d\vec{x} = -W \tag{3.72}$$

Note Well: that only one limit of integration depends on x; the other depends on where you choose to make the potential energy zero. This is a *free choice*. No physical result that can be measured or observed can uniquely depend on where you choose the potential energy to be zero. Let's understand this.

3.4.1: Force from Potential Energy

In one dimension, the x-component of $-\vec{F} \cdot d\vec{\ell}$ is:

$$dU = -dW = -F_x dx \tag{3.73}$$

If we rearrange this, we get:

$$F_x = -\frac{dU}{dx} \tag{3.74}$$

That is, the force is the *slope of the potential energy function*. This is actually a rather profound result and relationship.

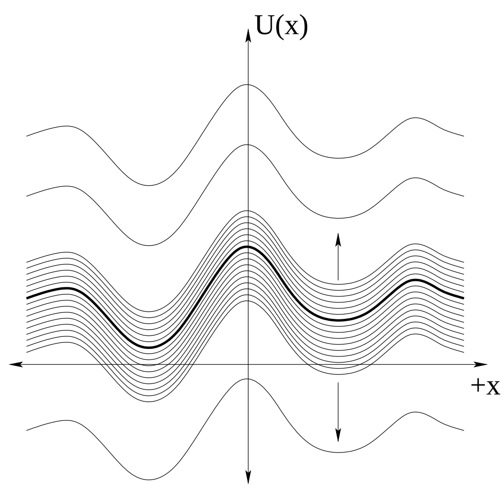


Figure 3.8: A tiny subset of the infinite number of possible $U(x)$ functions that might lead to the same physical force $F_x(x)$. One of these is highlighted by means of a thick line, but the only thing that might make it "preferred" is whether or not it makes solving any given problem a bit easier.

Consider the set of transformations that leave the *slope* of a function invariant. One of them is quite obvious – adding a positive or negative constant to $U(x)$ as portrayed in figure 3.8 does not affect its slope with respect to x, it just moves the whole function up or down on the U-axis. That means that all of this infinite set of candidate potential energies that differ by only a constant overall energy *lead to the same force!*

That's *good*, as force is something we can often measure, even "at a point" (without necessarily moving the object), but potential energy is *not*. To measure the work done by a conservative force on an object (and hence measure the change in the potential energy) we have to permit the force to *move* the object from one place to another and measure the change in its speed, hence its kinetic energy. We only measure a *change*, though – we cannot directly measure the absolute magnitude of the potential energy, any more than we can point to an object and say that the work of that object is zero Joules, or ten Joules, or whatever. We can talk about the amount of work done *moving* the object from *here* to *there* but objects do not possess "work" as an attribute, and potential energy is *just a convenient renaming of the work*, at least so far.

I cannot, then, tell you precisely what the near-Earth gravitational potential energy of a 1 kilogram mass sitting on a table is, not even if you tell me *exactly* where the table and the mass are in some sort of Universal coordinate system (where if the latter exists, as now seems dubious given our discussion of inertial frames and so on, we have yet to find it). There are literally an infinity of possible answers that will all *equally well predict the outcome of any physical experiment* involving near-Earth gravity acting on the mass, because they all lead to the same *force* acting on the object.

In more than one dimesion we have to use a bit of vector calculus to write this same result *per component*:

$$\Delta U = -\int \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{\ell}} \tag{3.75}$$

$$dU = -\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{\ell}} \tag{3.76}$$

It's a bit more work than we can do in this course to prove it, but the result one gets by "dividing through but $d\vec{\boldsymbol{\ell}}$" in this case is:

$$\vec{\boldsymbol{F}} = -\vec{\boldsymbol{\nabla}} U = -\frac{\partial U}{\partial x}\hat{\boldsymbol{x}} - \frac{\partial U}{\partial y}\hat{\boldsymbol{y}} - \frac{\partial U}{\partial z}\hat{\boldsymbol{z}} \tag{3.77}$$

or, the vector force is the negative ***gradient*** of the potential energy. This is basically the one dimensional result written above, per component.

If you are a physics or math major (or have had or are in multivariate calculus) so that you know what a gradient *is*, this last form should probably be studied until it makes sense, but *everybody* should know that

$$F_x = -\frac{dU}{dx} \tag{3.78}$$

in any given direction so this relation should reasonably hold (subject to working out some more math than you may yet know) for all coordinate directions. Note that non-physics majors won't (in my classes) be held *responsible* for knowing this vector calculus form, but everybody should understand the concept underlying it. We'll discuss this a bit further below, after we have learned about the total mechanical energy.

To summarize: we now know the definition of a conservative force, its potential energy, and how to get the force back from the potential energy. We hopefully are starting to understand our freedom to choose add a constant energy to the potential energy and still get the same answers to all physics problems[101] – obviously adding a constant to the potential energies won't change the *derivatives* of the potential energy function and hence won't lead to a different *force*.

Still, we had a perfectly good way of solving problems where we wished to find v as a function of x or vice versa (independent of time): the Work-Kinetic Energy Theorem. Why, then, do we bother inventing all of this complication: conservative forces, potential energies, arbitrary zeros? What was wrong with plain old work?

Well, for one thing, since the work done by conservative forces is independent of the path taken *by definition*, we can do the work integrals *once and for all* for the well-known conservative forces, stick a minus sign in front of them, and have a set of well-known ***potential energy functions*** that are generally even simpler and more useful. In fact, since one can easily differentiate the potential energy function to recover the force, one can in fact *forget thinking in terms of the force altogether* and formulate all of physics in terms of energies and potential energy functions!

In *this* class, we won't go to this extreme – we will simply learn **both** the forces **and** the associated potential energy functions where appropriate (there aren't that many; this isn't like learning all of organic chemistry's reaction pathways, for example...), deriving the second from the first as we go, but in future courses taken by a physics major, a chemistry major, a math major it is quite likely that you will relearn even classical mechanics in terms of the *Lagrangian*[102] or *Hamiltonian*[103] formulation, both of which are fundamentally energy-based, and quantum physics is almost entirely derived and understood in terms of Hamiltonians.

For now let's see how life is made a bit simpler by deriving general forms for the potential energy functions for near-Earth gravity and masses on springs, both of which will be very useful indeed to us in the weeks to come.

3.4.2: Potential Energy Function for Near-Earth Gravity

The potential energy of an object experiencing a near-Earth gravitational force is either:

$$U_g(y) = -\int_0^y (-mg)dy' = mgy \tag{3.79}$$

where we have effectively set the zero of the potential energy to be "ground level", at least if we put the y-coordinate origin at the ground. Of course, we don't really need to do this – we might well want the zero to be at the top of a table *over* the ground, or the top of a cliff well above that, and we are free to do so. More generally, we can write the gravitational potential

[101] Wikipedia: http://www.wikipedia.org/wiki/Gauge Theory. For students intending to continue with more physics, this is perhaps your first example of an idea called *Gauge freedom* – the invariance of things like energy under certain sets of coordinate transformations and the implications (like invariance of a measured force) of the symmetry groups of those transformations – which turns out to be very important indeed in future courses. And if this sounds strangely like I'm speaking Martian to you or talking about your freedom to choose a 12 gauge shotgun instead of a 20 gauge shotgun – gauge freedom indeed – well, don't worry about it...

[102] Wikipedia: http://www.wikipedia.org/wiki/Lagrangian.

[103] Wikipedia: http://www.wikipedia.org/wiki/Hamiltonian.

energy as the *indefinite* integral:

$$U_g(y) = -\int (-mg)dy = mgy + U_0 \tag{3.80}$$

where U_0 is an arbitrary constant that sets the zero of gravitational potential energy. For example, suppose we *did* want the potential energy to be zero at the top of a cliff of height H, but for one reason or another selected a coordinate system with the y-origin at the bottom. Then we need:

$$U_g(y = H) = mgH + U_0 = 0 \tag{3.81}$$

or

$$U_0 = -mgH \tag{3.82}$$

so that:

$$U_g(y) = mgy - mgH = mg(y - H) = mgy' \tag{3.83}$$

where in the last step we *changed variables* (coordinate systems) to a new one $y' = y - H$ with the origin at the top of the cliff!

From the latter, we see that our freedom to choose any location for the zero of our potential energy function is somehow tied to our freedom to choose an arbitrary origin for our coordinate frame. It is actually even more powerful (and more general) than that – we will see examples later where potential energy can be defined to be zero on entire planes or lines or "at infinity", where of course it is difficult to put an origin at infinity and have local coordinates make any sense.

You will find it **very helpful** to choose a coordinate system *and* set the zero of potential energy in such as way as to make the problem as computationally simple as possible. Only experience and practice will ultimately be your best guide as to just what those are likely to be.

3.4.3: Springs

Springs also exert conservative forces in one dimension – the work you do compressing or stretching an ideal spring equals the work the spring does going back to its original position, whatever that position might be. We can therefore define a potential energy function for them.

In most cases, we will choose the zero of potential energy to be the *equilibrium position of the spring* – other choices are possible, though, and one in particular will be useful (a mass hanging from a spring in near-Earth gravity).

With the zero of both our one dimensional coordinate system and the potential energy at the equilibrium position of the unstretched spring (easiest) Hooke's Law is just:

$$F_x = -kx \tag{3.84}$$

and we get:

$$\begin{aligned} U_s(x) &= -\int_0^x (-kx')\,dx' \\ &= \frac{1}{2}kx^2 \end{aligned} \tag{3.85}$$

This is the function you should learn – by deriving this result several times on your own, not by memorizing – as the potential energy of a spring.

More generally, if we do the indefinite integral in this coordinate frame instead we get:

$$U(x) = -\int (-kx)\ dx = \frac{1}{2}kx^2 + U_0 \tag{3.86}$$

To see how this is related to one's choice of coordinate origin, suppose we choose the origin of coordinates to be at the end of the spring fixed to a wall, so that the equilibrium length of the unstretched, uncompressed spring is $x_{\rm eq}$. Hooke's Law is written in *these* coordinates as:

$$F_x(x) = -k(x - x_{\rm eq}) \tag{3.87}$$

Now we can choose the zero of potential energy to be at the position $x = 0$ by doing the definite integral:

$$U_s(x) = -\int_0^x \left(-k(x' - x_{\rm eq})\right)\ dx' = \frac{1}{2}k(x - x_{\rm eq})^2 - \frac{1}{2}kx_{\rm eq}^2 \tag{3.88}$$

If we now change variables to, say, $y = x - x_{\rm eq}$, this is just:

$$U_s(y) = \frac{1}{2}ky^2 - \frac{1}{2}kx_{\rm eq}^2 = \frac{1}{2}ky^2 + U_0 \tag{3.89}$$

which can be compared to the indefinite integral form above. Later, we'll do a problem where a mass hangs from a spring and see that our freedom to add an arbitrary constant of integration allows us to change variables to an "easier" origin of coordinates halfway through a problem.

Consider: our treatment of the spring gun (above) would have been simpler, would it not, if we could have simply started knowing the potential energy function for (and hence the work done by) a spring?

There is one more way that using potential energy instead of work per se will turn out to be useful to us, and it is the motivation for including the leading minus sign in its definition. Suppose that you have a mass m that is moving under the influence of a *conservative force*. Then the Work-Kinetic Energy Theorem (3.35) looks like:

$$W_C = \Delta K \tag{3.90}$$

where W_C is the ordinary work done by the conservative force. Subtracting W_C over to the other side and substituting, one gets:

$$\Delta K - W_C = \Delta K + \Delta U = 0 \tag{3.91}$$

Since we can now assign $U(\vec{x})$ a *unique value* (once we set the constant of integration or place(s) $U(\vec{x})$ is zero in its definition above) at each point in space, and since K is similarly a function of position in space when time is eliminated in favor of position and no other (non-conservative) forces are acting, we can define the *total mechanical energy* of the particle to be:

$$E_{\rm mech} = K + U \tag{3.92}$$

in which case we just showed that

$$\Delta E_{\rm mech} = 0 \tag{3.93}$$

Wait, did we just prove that $E_{\rm mech}$ is a constant any time a particle moves around under only the influence of conservative forces? We did...

3.5: Conservation of Mechanical Energy

OK, so maybe you missed that last little bit. Let's make it a bit clearer and see how *enormously* useful and important this idea is.

First we will state the principle of the ***Conservation of Mechanical Energy***:

> **The total mechanical energy (defined as the sum of its potential and kinetic energies) of a particle being acted on by only conservative forces is constant.**

Or (in much more concise algebra), if only conservative forces act on an object and U is the potential energy function for the total conservative force, then

$$E_{\rm mech} = K + U = \text{A scalar constant} \tag{3.94}$$

The proof of this statement is given above, but we can recapitulate it here.

Suppose

$$E_{\rm mech} = K + U \tag{3.95}$$

Because the change in potential energy of an object is just the path-independent negative work done by the conservative force,

$$\Delta K + \Delta U = \Delta K - W_C = 0 \tag{3.96}$$

is just a restatement of the WKE Theorem, which we *derived and proved*. So it must be true! But then

$$\Delta K + \Delta U = \Delta(K + U) = \Delta E_{\rm mech} = 0 \tag{3.97}$$

and $E_{\rm mech}$ must be constant as the conservative force moves the mass(es) around.

3.5.1: Force, Potential Energy, and Total Mechanical Energy

Now that we know what the total mechanical energy is, the following little litany might help you conceptually grasp the relationship between potential energy and force. We will return to this still again below, when we talk about potential energy curves and equilibrium, but repetition makes the ideas easier to understand and remember, so skim it here first, now.

The fact that the force is the *negative* derivative of (or gradient of) the potential energy of an object ***means*** that ***the force points in the direction the potential energy decreases in***.

This makes *sense*. If the object has a constant total energy, and it moves *in* the direction of the force, it speeds up! Its kinetic energy increases, therefore its potential energy decreases. If it moves from lower potential energy to higher potential energy, its kinetic energy decreases, which means the force pointed the other way, slowing it down.

There is a simple metaphor for all of this – the slope of a hill. We all know that things roll slowly down a shallow hill, rapidly down a steep hill, and just fall right off of cliffs. The force that speeds them up is related to the *slope* of the hill, and so is the rate at which their gravitational potential energy increases as one goes down the slope! In fact, it isn't actually just a metaphor, more like an example.

Either way, "downhill" is where potential energy variations push objects – in the direction that the potential energy maximally decreases, with a force proportional to the rate at which it decreases. The WKE Theorem itself and all of our results in this chapter, after all, are derived from Newton's Second Law – energy conservation is just Newton's Second Law in a time-independent disguise.

Example 3.5.1: Falling Ball Reprise

To see how powerful this is, let us look back at a falling object of mass m (neglecting drag and friction). First, we have to determine the gravitational potential energy of the object a height y above the ground (where we will choose to set $U(0) = 0$):

$$U(y) = -\int_0^y (-mg)dy = mgy \tag{3.98}$$

Wow, that was kind of – easy!

Now, suppose we have our ball of mass m at the height H and drop it from rest, yadda yadda. How fast is it going when it hits the ground? *This* time we simply write the total energy of the ball at the top (where the potential is mgH and the kinetic is zero) and the bottom (where the potential is zero and kinetic is $\frac{1}{2}mv^2$) and set the two equal! Solve for v, done:

$$E_i = mgH + (0) = (0) + \frac{1}{2}mv^2 = E_f \tag{3.99}$$

or

$$v = \sqrt{2gH} \tag{3.100}$$

Even better:

Example 3.5.2: Block Sliding Down Frictionless Incline Reprise

The block starts out a height H above the ground, with potential energy mgH and kinetic energy of 0. It slides to the ground (no non-conservative friction!) and arrives with no potential energy and kinetic energy $\frac{1}{2}mv^2$. Whoops, time to block-copy the previous solution:

$$E_i = mgH + (0) = (0) + \frac{1}{2}mv^2 = E_f \tag{3.101}$$

or

$$v = \sqrt{2gH} \tag{3.102}$$

Example 3.5.3: A Simple Pendulum

Here are two versions of a pendulum problem: Imagine a pendulum (ball of mass m suspended on a string of length L that we have pulled up so that the ball is a height $H < L$ above its lowest point on the arc of its stretched string motion. We release it from rest. How fast is it going at the bottom? Yep, you guessed it – block copy again:

$$E_i = mgH + (0) = (0) + \frac{1}{2}mv^2 = E_f \tag{3.103}$$

or

$$v = \sqrt{2gH} \tag{3.104}$$

It looks as though *it does not matter* what path a mass takes as it goes down a height H starting from rest – as long as no forces act to *dissipate* or *add* energy to the particle, it will arrive at the bottom travelling at the same speed.

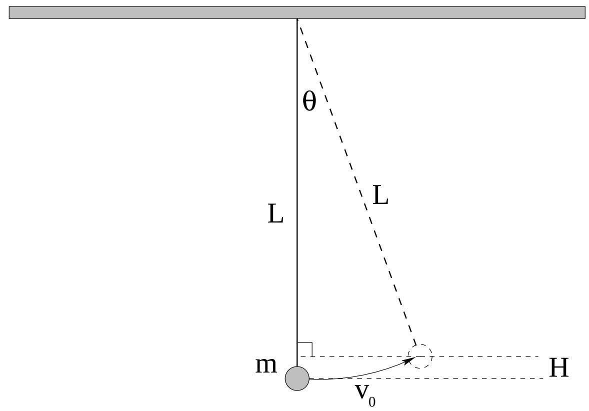


Figure 3.9: Find the maximum angle through which the pendulum swings from the initial conditions.

Here's the same problem, formulated a different way: A mass m is hanging by a massless thread of length L and is given an initial speed v_0 to the right (at the bottom). It swings up and stops at some maximum height H at an angle θ as illustrated in figure 3.9 (which can be used "backwards" as the figure for the first part of this example, of course). Find θ.

Again we solve this by setting $E_i = E_f$ (total energy is conserved).

Initial:

$$E_i = \frac{1}{2}mv_0^2 + mg(0) = \frac{1}{2}mv_0^2 \tag{3.105}$$

Final:

$$E_f = \frac{1}{2}m(0)^2 + mgH = mgL(1 - \cos(\theta)) \tag{3.106}$$

(Note well: $H = L(1 - \cos(\theta))$!)

Set them equal and solve:

$$\cos(\theta) = 1 - \frac{v_0^2}{2gL} \tag{3.107}$$

or

$$\theta = \cos^{-1}(1 - \frac{v_0^2}{2gL}). \tag{3.108}$$

Example 3.5.4: Looping the Loop

Here is a lovely problem – so lovely that you will solve it five or six times, at least, in various forms throughout the semester, so be sure that you get to where you understand it – that requires you to use *three different principles* we've learned so far to solve:

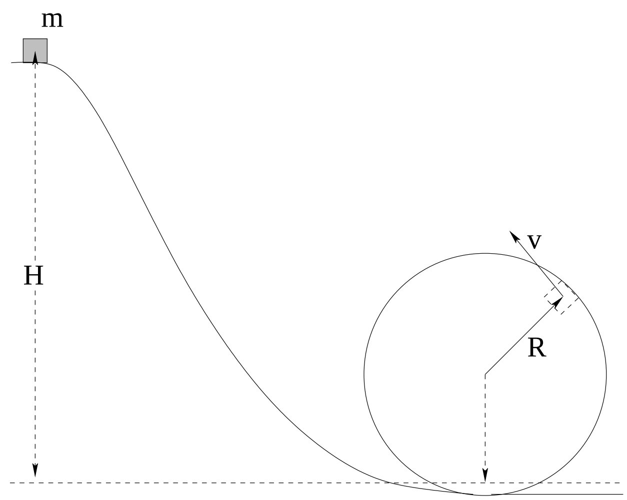


Figure 3.10

What is the minimum height H such that a block of mass m loops-the-loop (stays ***on the frictionless track all the way around the circle***) in figure 3.10 above?

Such a simple problem, such an involved answer. Here's how you might proceed. First of all, let's understand the condition that must be satisfied for the answer "stay on the track". For a block to stay on the track, it has to *touch* the track, and touching something means "exerting a normal force on it" in physicsspeak. To *barely* stay on the track, then – the minimal condition – is for the normal force to *barely* go to zero.

Fine, so we need the block to precisely "kiss" the track at near-zero normal force at the point where we expect the normal force to be weakest. And where is that? Well, at the place it is moving the *slowest*, that is to say, the *top of the loop*. If it comes off of the loop, it is bound to come off at or before it reaches the top.

Why is that point key, and *what is the normal force doing* in this problem. Here we need two physical principles: ***Newton's Second Law*** and ***the kinematics of circular motion*** since the mass is undoubtedly moving in a circle if it stays on the track. Here's the way we reason: "If the block moves in a circle of radius R at speed v, then its acceleration towards the center must be $a_c = v^2/R$. Newton's Second Law then tells us that the *total force component* in the direction of the center must be mv^2/R. That force can only be made out of (a component of) gravity and the normal force, which points towards the center. So we can relate the normal force to the speed of the block on the circle at any point."

At the top (where we expect v to be at its minimum value, assuming it stays on the circle) gravity points straight towards the center of the circle of motion, so we get:

$$mg + N = \frac{mv^2}{R} \tag{3.109}$$

and in the limit that $N \to 0$ ("barely" looping the loop) we get the condition:

$$mg = \frac{mv_t^2}{R} \tag{3.110}$$

where v_t is the (minimum) speed at the top of the track needed to loop the loop.

Now we need to relate the speed at the top of the circle to the original height H it began at. This is where we need our third principle – *Conservation of Mechanical Energy!* Note that we cannot possible integrate Newton's Second Law and solve an equation of motion for the block on the frictionless track – I haven't given you any sort of equation for the track (because I don't know it) and even a physics graduate student forced to integrate N2 to find the answer for some relatively "simple" functional form for the track would suffer mightily finding the answer. With energy we don't care about the shape of the track, only that the track do no work on the mass which (since it is frictionless and normal forces do no work) is in the bag.

Thus:

$$E_i = mgH = mg2R + \frac{1}{2}mv_t^2 = E_f \tag{3.111}$$

If you put these two equations together (e.g. solve the first for mv_t^2 and substitute it into the second, then solve for H in terms of R) you should get $H_{\min} = 5R/2$. Give it a try. You'll get even more practice in your homework, for some more complicated situations, for masses on strings or rods – they're all the same problem, but sometimes the Newton's Law condition will be quite different! Use your intuition and experience with the world to help guide you to the right solution in all of these causes.

So *any time* a mass moves down a distance H, its change in potential energy is mgH, and since total mechanical energy is conserved, its change in kinetic energy is *also* mgH the other way. As one increases, the other decreases, and vice versa!

This makes kinetic and potential energy bone simple to use. It also means that there is a lovely analogy between potential energy and your savings account, kinetic energy and your checking account, and cash transfers (conservative movement of money from checking to savings or vice versa where your *total* account remains constant.

Of course, it is almost too much to expect for life to *really* be like that. We know that we always have to pay banking fees, teller fees, taxes, somehow we *never* can move money around and end up with as much as we started with. And so it is with energy.

Well, it is and it isn't. Actually conservation of energy is a very deep and fundamental principle of the entire Universe as best we can tell. Energy seems to be conserved everywhere, all of the time, in detail, to the best of our ability to experimentally check. However, *useful* energy tends to decrease over time because of "taxes". The tax collectors, as it were, of nature are *non-conservative forces*!

What happens when we try to combine the work done by non-conservative forces (which we must tediously calculate per problem, per path) with the work done by conservative ones, expressed in terms of potential and total mechanical energy? You get the...

3.6: Generalized Work-Mechanical Energy Theorem

So, as suggested above let's generalize the WKE one further step by considering what happens if ***both*** conservative and nonconservative forces are acting ***on a particle***. In that case the argument above becomes:

$$W_{\text{tot}} = W_C + W_{NC} = \Delta K \tag{3.112}$$

or

$$W_{NC} = \Delta K - W_C = \Delta K + \Delta U = \Delta E_{\rm mech} \tag{3.113}$$

which we state as the **Generalized Non-Conservative Work-Mechanical Energy Theorem**:

> **The work done by all the non-conservative forces acting on a particle equals the change in its total mechanical energy.**

Our example here is very simple.

Example 3.6.1: Block Sliding Down a Rough Incline

Suppose a block of mass m slides down an incline of length L at an incline θ with respect to the horizontal and with kinetic friction (coefficient μ_k) acting against gravity. How fast is it going (released from rest at an angle where static friction cannot hold it) when it reaches the ground?

Here we have to do a mixture of several things. First, let's write Newton's Second Law for just the (static) y direction:

$$N - mg\cos(\theta) = 0 \tag{3.114}$$

or

$$N = mg\cos(\theta) \tag{3.115}$$

Next, evaluate:

$$f_k = \mu_k N = \mu_k mg\cos(\theta) \tag{3.116}$$

(up the incline, opposite to the motion of the block).

We *ignore* dynamics in the direction down the plane. Instead, we note that the work done by friction is equal to the change in the mechanical energy of the block. $E_i = mgH = mgL\sin(\theta)$. $E_f = \frac{1}{2}mv^2$. So:

$$-\mu_k mgL\cos(\theta) = E_f - E_i = \frac{1}{2}mv^2 - mgH \tag{3.117}$$

or

$$\frac{1}{2}mv^2 = mgH - \mu_k mgL\cos(\theta) \tag{3.118}$$

so that:

$$v = \pm\sqrt{2gH - \mu_k 2gL\cos(\theta)} \tag{3.119}$$

Here we really do have to be careful and choose the sign that means "going down the incline" at the bottom.

As an extra bonus, our answer tells us the condition on (say) the angle such that the mass doesn't or just barely *makes* it to the bottom. $v = 0$ means "barely" (gets there and stops) and if v is *imaginary*, it doesn't make all the way to the bottom at all.

I don't know about you, but this seems a lot easier than messing with integrating Newton's Law, solving for $v(t)$ and $x(t)$, solving for t, back substituting, etc. It's not that this is all that difficult, but work-energy is simple bookkeeping, anybody can do it if they just know stuff like the form of the potential energy, the magnitude of the force, some simple integrals.

Here's another example.

Example 3.6.2: A Spring and Rough Incline

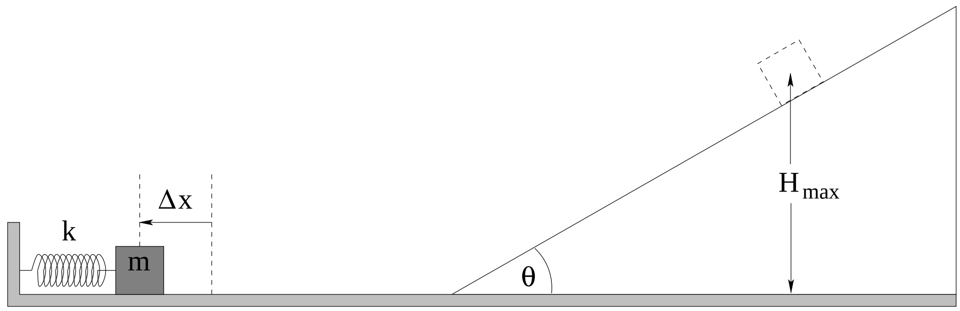


Figure 3.11: A spring compressed an initial distance Δx fires a mass m across a smooth ($\mu_k \approx 0$) floor to rise up a rough $\mu_k \neq 0$) incline. How far up the incline does it travel before coming to rest?

In figure 3.11 a mass m is released from rest from a position on a spring with spring constant k compressed a distance Δx from equilibrium. It slides down a frictionless horizontal surface and then slides up a rough plane inclined at an angle θ. What is the maximum height that it reaches on the incline?

This is a problem that is basically impossible, so far, for us to do using Newton's Laws alone. This is because we are weeks away from being able to solve the equation of motion for the mass on the spring! Even if/when we *can* solve the equation of motion for the mass on the spring, though, this problem would still be quite painful to solve using Newton's Laws and dynamical solutions directly.

Using the GWME Theorem, though, it is pretty easy. As before, we have to express the initial total mechanical energy and the final total mechanical energy algebraically, and set their *difference* equal to the non-conservative work done by the force of kinetic friction sliding up the incline.

I'm not going to do *every step* for you, as this seems like it would be a good homework problem, but here are a few:

$$E_i = U_{gi} + U_{si} + K_i = mg(0) + \frac{1}{2}k\Delta x^2 + \frac{1}{2}m(0)^2 = \frac{1}{2}k\Delta x^2 \tag{3.120}$$

$$E_f = U_{gf} + K_f = mgH_{\max} + \frac{1}{2}m(0)^2 = mgH_{\max} \tag{3.121}$$

Remaining for you to do: Find the force of friction down the incline (as it slides up). Find the work done by friction. Relate that work to $H_{\max}$ algebraically, write the GWME Theorem algebraically, and solve for $H_{\max}$. Most of the steps involving friction and the inclined plane can be found in the previous example, if you get stuck, but try to do it without looking first!

3.6.1: Heat and Conservation of Energy

Note well that the theorem above only applies to forces acting on *particles*, or objects that we consider in the particle approximation (ignoring any internal structure and treating the object like a single mass). In fact, all of the rules above (so far) from Newton's Laws on down strictly speaking only apply to particles in inertial reference frames, and we have some work to do in order to figure out how to apply them to *systems* of particles being pushed on both by *internal* forces between particles in the system as well as *external* forces between the particles of the system and particles that are not part of the system.

What happens to the energy added to or removed from an object (that is really made up of *many* particles bound together by internal e.g. molecular forces) by things like my non-conservative hand as I give a block treated as a "particle" a push, or non-conservative kinetic friction and drag forces as they act on the block to slow it down as it slides along a table? This is *not* a trivial question. To properly answer it we have to descend all the way into the conceptual abyss of treating every single particle that makes up the system we call "the block" and every single particle that makes up the system consisting of "everything else in the Universe but the block" and all of the internal forces between them – which happen, as far as we can tell, to be *strictly conservative forces* – and then somehow average over them to recover the ability to treat the block like a particle, the table like a fixed, immovable object it slides on, and friction like a comparatively simple force that does non-conservative work on the block.

It requires us to invent things like statistical mechanics to do the averaging, thermodynamics to describe certain kinds of averaged systems, and whole new sciences such as chemistry and biology that use averaged energy concepts with their own fairly stable rules that cannot *easily* be connected back to the microscopic interactions that bind quarks and electrons into atoms and atoms together into molecules. It's easy to get lost in this, because it is both fascinating and *really difficult*.

I'm therefore going to give you a very important empirical law (that we can understand well enough from our treatment of particles so far) and a rather *heuristic* description of the connections between microscopic interactions and energy and the macroscopic mechanical energy of things like blocks, or cars, or human bodies.

The important empirical law is the ***Law of Conservation of Energy***[104]. Whenever we examine a physical system and try very hard to keep track of all of the mechanical energy exchanges withing that system and between the system and its surroundings, we find that we can always account for them all without any gain or loss. In other words, we find that the total mechanical energy of an ***isolated*** system never changes, and if we add or remove mechanical energy to/from the system, it has to come from or go to somewhere outside of the system. This result, applied to well defined systems of particles, can be formulated as the ***First Law of Thermodynamics***:

$$\Delta Q_{\rm in} = \Delta E_{\rm of} + W_{\rm by} \tag{3.122}$$

In words, the heat energy flowing *in* to a system equals the change in the internal total mechanical energy *of* the system plus the external work (if any) done *by* the system on its surroundings. The total mechanical energy of the system itself is just the sum of the potential and kinetic energies of all of its internal parts and is simple enough to understand if not to compute. The work done by the system on its surroundings is similarly simple enough to understand if not to compute. The hard part of this law is the definition of *heat energy*, and sadly, I'm not going to give you more than the crudest idea of it right now and make some

[104]More properly, mass-energy, but we really don't want to get into that in an introductory course.

statements that aren't strictly true because to treat heat *correctly* requires a major chunk of a whole new textbook on textbfThermodynamics. So take the following with a grain of salt, so to speak.

When a block slides down a rough table from some initial velocity to rest, kinetic friction turns the bulk organized kinetic energy of the *collectively* moving mass into ***disorganized microscopic energy*** – heat. As the rough microscopic surfaces bounce off of one another and form and break chemical bonds, it sets the actual molecules of the block bounding, increasing the internal microscopic mechanical energy of the block and *warming it up*. Some of it similarly increasing the internal microscopic mechanical energy of the table it slide across, warming *it* up. Some of it appears as light energy (electromagnetic radiation) or sound energy – initially organized energy forms that themselves become ever more disorganized. Eventually, the initial organized energy of the block becomes a tiny increase in the average internal mechanical energy of a very, very large number of objects both inside and outside of the original system that we call the block, a process we call being "lost to heat".

We have the same sort of problem tracking energy that we *add* to the system when I give the block a push. Chemical energy in sugars causes muscle cells to change their shape, contracting muscles that do work on my arm, which exchanges energy with the block via the normal force between block and skin. The chemical energy itself originally came from thermonuclear fusion reactions in the sun, and the free energy released in those interactions can be tracked back to the Big Bang, with a lot of imagination and sloughing over of details. Energy, it turns out, has "always" been around (as far back in time as we can see, literally) but is constantly changing form and generally becoming more disorganized as it does so.

In this textbook, we will say a little more on this later, but this is enough for the moment. We will summarize this discussion by remarking that non-conservative forces, both external (e.g. friction acting on a block) and internal (e.g. friction or collision forces acting between two bodies that are part of "the system" being considered) will often do work that entirely or partially "turns into heat" – disappears from the total mechanical energy we can easily track. That doesn't mean that it is has truly disappeared, and more complex treatments or experiments can indeed track and/or measure it, but we just barely learned what mechanical energy *is* and are not yet ready to try to deal with what happens when it is shared among (say) Avogadro's number of interacting gas molecules.

3.7: Power

The energy in a given system is **not**, of course, usually constant in time. Energy is added to a given mass, or taken away, at some rate. We accelerate a car (adding to its mechanical energy). We brake a car (turning its kinetic energy into heat). There are many times when we are given the **rate** at which energy is added or removed in time, and need to find the total energy added or removed. This rate is called the **power**.

Power: The rate at which work is done, or energy released into a system.

This latter form lets us express it conveniently for time-varying forces:

$$dW = \vec{F} \cdot d\vec{x} = \vec{F} \cdot \frac{dx}{dt} dt \tag{3.123}$$

or

$$P = \frac{dW}{dt} = \vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{v}} \tag{3.124}$$

so that

$$\Delta W = \Delta E_{\text{tot}} = \int P dt \tag{3.125}$$

The units of power are clearly Joules/sec = Watts. Another common unit of power is "Horsepower", 1 HP = 746 W. Note that the power of a car together with its drag coefficient determine how fast it can go. When energy is being added by the engine at the same **rate** at which it is being dissipated by drag and friction, the total mechanical energy of the car will remain constant in time.

Example 3.7.1: Rocket Power

A model rocket engine delivers a constant thrust F that pushes the rocket (of approximately constant mass m) up for a time t_r before shutting off. Show that the total energy delivered by the rocket engine is equal to the change in mechanical energy *the hard way* – by solving Newton's Second Law for the rocket to obtain $v(t)$, using that to find the power P, and integrating the power from 0 to t_r to find the total work done by the rocket engine, and comparing this to $mgy(t_r) + \frac{1}{2}mv(t_r)^2$, the total mechanical energy of the rocket at time t_r.

To outline the solution, following a previous homework problem, we write:

$$F - mg = ma \tag{3.126}$$

or

$$a = \frac{F - mg}{m} \tag{3.127}$$

We integrate twice to obtain (starting at $y(0) = 0$ and $v(0) = 0$):

$$\begin{aligned} v(t) &= at = \frac{F - mg}{m} t & (3.128) \\ y(t) &= \frac{1}{2}at^2 = \frac{1}{2}\frac{F - mg}{m} t^2 & (3.129) \\ & & (3.130) \end{aligned}$$

From this we can find:

$$\begin{aligned} E_{\text{mech}}(t_r) &= mgy(t_r) + \frac{1}{2}mv(t_r)^2 \\ &= mg\frac{1}{2}\frac{F - mg}{m}t_r^2 + \frac{1}{2}m\left(\frac{F - mg}{m}t_r\right)^2 \\ &= \frac{1}{2}\left(Fg - mg^2 + \frac{F^2}{m} + mg^2 - 2Fg\right)t_r^2 \\ &= \frac{1}{2m}\left(F^2 - Fmg\right)t_r^2 & (3.131) \end{aligned}$$

Now for the power:

$$\begin{aligned} P &= F \cdot v = Fv(t) \\ &= \frac{F^2 - Fmg}{m} t & (3.132) \end{aligned}$$

We integrate this from 0 to t_r to find the total energy delivered by the rocket engine:

$$W = \int_0^{t_r} P\, dt = \frac{1}{2m}\left(F^2 - Fmg\right) t_r^2 = E_{\rm mech}(t_r) \tag{3.133}$$

For what it is worth, this should *also* just be $W = F \times y(t_r)$, the force through the distance:

$$W = F \times \frac{1}{2}\frac{F - mg}{m} t_r^2 = \frac{1}{2m}\left(F^2 - Fmg\right) t_r^2 \tag{3.134}$$

and it is.

The main point of this example is to show that all of the definitions and calculus above are *consistent*. It doesn't matter how you proceed – compute $\Delta E_{\rm mech}$, find $P(t)$ and integrate, or just straight up evaluate the work $W = F\Delta y$, you will get the same answer.

Power is an extremely important quantity, *especially* for engines because (as you see) the faster you go at constant thrust, the larger the power delivery. Most engines have a limit on the amount of power they can generate. Consequently the *forward directed force or thrust* tends to fall off as the speed of the e.g. rocket or car increases. In the case of a car, the car must also overcome a (probably nonlinear!) drag force. One of your homework problems explores the economic consequences of this.

3.8: Equilibrium

Recall that the **force** is given by the **negative gradient of the potential energy**:

$$\vec{\boldsymbol{F}} = -\vec{\boldsymbol{\nabla}} U \tag{3.135}$$

or (in each direction[105]):

$$F_x = -\frac{dU}{dx}, \quad F_y = -\frac{dU}{dy}, \quad F_z = -\frac{dU}{dz} \tag{3.136}$$

or the force is the negative **slope** of the potential energy function in this direction. As discussed above, the *meaning* of this is that if a particle moves in the direction of the (conservative) force, it speeds up. If it speeds up, its kinetic energy increases. If its kinetic energy increases, its potential energy must *decrease*. The force (component) acting on a particle is thus the *rate at which the potential energy decreases* (the negative slope) in any given direction as shown.

In the discussion below, we will concentrate on one-dimensional potentials to avoid over-stressing students' calculus muscles while they are still under development, but the ideas all generalize beautifully to two or three (or in principle still more) dimensions.

In one dimension, we can use this to rapidly evaluate the behavior of a system on a qualitative basis just by *looking* at a *graph of the curve*! Consider the potential energy curves in figure 3.12. At the point labelled a, the x-slope of $U(x)$ is *positive*. The x (component of the) force, therefore, is in the negative x direction. At the point b, the x-slope is *negative* and the

[105]Again, more advanced math or physics students will note that these should all be *partial* derivatives in correspondance with the force being the *gradient* of the potential energy surface $U(x, y, z)$, but even then each component is the local slope along the selected direction.

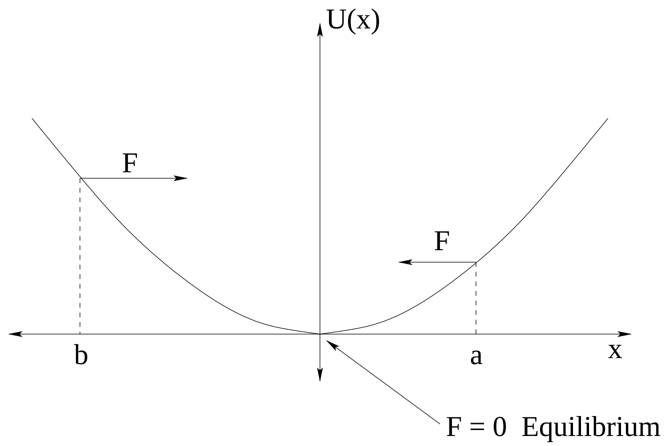


Figure 3.12: A one-dimensional potential energy curve $U(x)$. This particular curve might well represent $U(x) = \frac{1}{2}kx^2$ for a mass on a spring, but the features identified and classified below are generic.

force is correspondingly positive. Note well that the force gets larger as the slope of $U(x)$ gets larger (in magnitude).

The point in the middle, at $x = 0$, is *special*. Note that this is a *minimum* of $U(x)$ and hence the x-slope is zero. Therefore the x-directed force F at that point is zero as well. A point at which the force on an object is zero is, as we previously noted, a point of ***static force equilibrium*** – a particle placed there at rest will remain there at rest.

In this particular figure, if one moves the particle a small distance to the right or the left of the equilibrium point, the force *pushes the particle back towards equilibrium*. Points where the force is zero and small displacements cause a restoring force in this way are called *stable equilibrium points*. As you can see, the ***isolated minima*** of a potential energy curve (or surface, in higher dimensions) are all ***stable equilibria***.

Figure 3.13 corresponds to a more useful "generic" atomic or molecular interaction potential energy. It corresponds roughly to a *Van der Waals Force*[106] between two atoms or molecules, and exhibits a number of the features that such interactions often have.

At very long ranges, the forces between neutral atoms are extremely small, effectively zero. This is illustrated as an extended region where the potential energy is *flat* for large r. Such a range is called ***neutral equilibrium*** because there are no forces that either restore or repel the two atoms. Neutral equilibrium is *not stable* in the specific sense that a particle placed there with *any non-zero velocity* will move freely (according to Newton's First Law). Since it is nearly impossible to prepare an atom at absolute rest relative to another particle, one basically "never" sees two unbound microscopic atoms with a large, perfectly constant spatial orientation.

As the two atoms near one another, their interaction becomes first *weakly attractive* due to e.g. quantum dipole-induced dipole interactions and then *weakly repulsive* as the two atoms start to "touch" each other. There is a potential energy minimum in between where two atoms separated by a certain distance can be in stable equilibrium without being chemically bound.

[106]Wikipedia: http://www.wikipedia.org/wiki/Van der Waals Force.

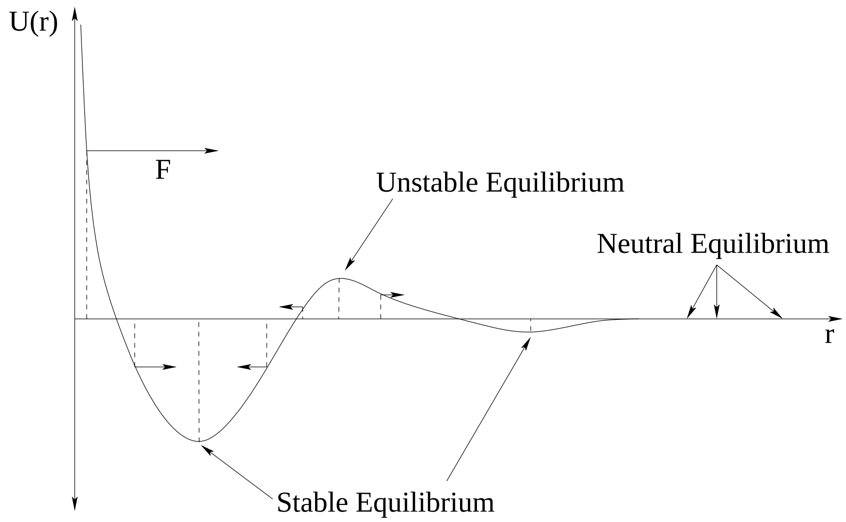


Figure 3.13: A fairly generic potential energy shape for microscopic (atomic or molecular) interactions, drawn to help exhibits *features* one might see in such a curve more than as a realistically scaled potential energy in some set of units. In particular, the curve exhibits stable, unstable, and neutral equilibria for a *radial* potential energy as a fuction of r, the distance between two e.g. atoms.

Atoms that approach one another still more closely encounter a *second* potential energy well that is at first *strongly* attractive (corresponding, if you like, to an actual chemical interaction between them) followed by a *hard core repulsion* as the electron clouds are prevented from interpenetrating by e.g. the Pauli exclusion principle. This second potential energy well is often modelled by a *Lennard-Jones potential energy* (or "6-12 potential energy", corresponding to the inverse powers of r used in the model[107] . It also has a point of stable equilibrium.

In between, there is a point where the growing attraction of the inner potential energy well and the growing repulsion of the outer potential energy well *balance*, so that the potential energy function has a *maximum*. At this maximum the slope is zero (so it is a position of force equilibrium) but because the force on either side of this point pushes the particle *away* from it, this is a point of ***unstable equilibrium***. Unstable equilibria occur at ***isolated maxima*** in the potential energy function, just as stable equilibria occur at *isolated minima*.

Note for advanced students: In more than one dimension, a potential energy curve can have "saddle points" that are maxima in one dimension and minima in another (so called because the potential energy surface resembles the surface of a saddle, curved up front-to-back to hold the rider in and curved down side to side to allow the legs to straddle the horse). Saddle points are *unstable* equilibria (because instability in any direction means unstable) and are of some conceptual importance in more advanced studies of physics or in mathematics when considering asymptotic convergence.

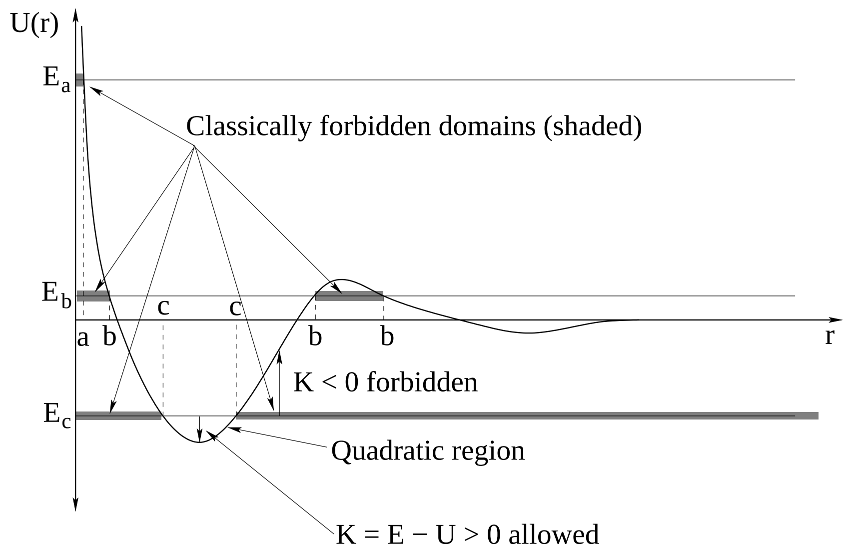


Figure 3.14: The same potential energy curve, this time used to illustrate *turning points* and classically *allowed and forbidden regions*. Understanding the role of the total energy on potential energy diagrams and how *transitions* from a higher energy state to a lower energy state can "bind" a system provide insight into chemistry, orbital dynamics, and more.

3.8.1: Energy Diagrams: Turning Points and Forbidden Regions

We now turn to another set of extremely useful information one can extract from potential energy curves in cases where one knows the *total mechanical energy* of the particle in addition to the potential energy curve. In figure 3.14 we again see the generic (Van der Waals) atomic interaction curve this time rather "decorated" with information. To understand this information and how to look at the diagram and gain insight, please read the following description very carefully while following along in the figure.

Consider a particle with total energy mechanical E_a. Since the total mechanical energy is a constant, we can draw the energy in on the potential energy axes as a *straight line with zero slope* – the same value for all r. Now note carefully that:

$$K(r) = E_{\rm mech} - U(r) \tag{3.137}$$

which is the *difference* between the total energy curve and the potential energy curve. The kinetic energy of a particle is $\frac{1}{2}mv^2$ which is non-negative. This means that we can *never* observe a particle with energy E_a to the left of the position marked a on the r-axis – only point where $E_a \geq U$ lead to $K > 0$. We refer to the point a as a ***turning point*** of the motion for any given energy – when $r = a$, $E_{\rm mech} = E_a = U(a)$ and $K(a) = 0$.

We can interpret the motion associated with E_a very easily. An atom comes in at more or less a constant speed from large r, speeds and slows and speeds again as it reaches the support of the potential energy[108], "collides" with the central atom at $r = 0$ (strongly repelled

[107]Wikipedia: http://www.wikipedia.org/wiki/Lennard-Jones Potential. We will learn the difference between a "potential energy" and a "potential" later in this course, but for the moment it is not important. The shapes of the two curves are effectively identical.

[108]The "support" of a function is the set values of the argument for which the function is not zero, in this case a finite sphere around the atom out where the potential energy first becomes attractive.

by the hard core interaction) and recoils, eventually receding from the central atom at more or less the same speed it initially came in with. Its distance of closest approach is $r = a$.

Now consider a particle coming in with energy $E_{\rm mech} = E_b$. Again, this is a constant straight line on the potential energy axes. Again $K(r) = E_b - U(r) \geq 0$. The points on the r-axis labelled b are the turning points of the motion, where $K(b) = 0$. The shaded regions indicate ***classically forbidden regions*** where ***the kinetic energy would have to be negative*** for the particle (with the given total energy) to be found there. Since the kinetic energy can never be negative, the atom can never be found there.

Again we can visualize the motion, but now there are two possibilities. If the atom comes in from infinity as before, it will initially be weakly attracted ultimately be slowed and repelled not by the hard core, but by the much softer force outside of the unstable maximum in $U(r)$. This sort of "soft" collision is an example of an ***interaction barrier*** a chemical reaction that cannot occur at low temperatures (where the energy of approach is too low to overcome this initial repulsion and allow the atoms to get close enough to chemically interact.

However, a second possibility emerges. If the separation of the two atoms (with energy E_b, recall) is in the classically allowed region between the two inner turning points, then the atoms will *oscillate* between those two points, unable to separate to infinity without passing through the classically forbidden region that would require the kinetic energy to be negative. The atoms in this case are said to be *bound* in a *classically stable* configuration around the stable equilibrium point associated with this well.

In nature, this configuration is generally *not* stably bound with an energy $E_b > 0$ – quantum theory permits an atom outside with this energy to *tunnel* into the inner well and an atom in the inner well to *tunnel* back to the outside and thence be repelled to $r \to \infty$. Atoms bound in this inner well are then said to be ***metastable*** (which means basically "slowly unstable") – they are classically bound for a while but eventually escape to infinity.

However, in nature pairs of atoms in the metastable configuration have a chance of *giving up some energy* (by, for example, giving up a photon or phonon, where you shouldn't worry too much about what these are just yet) and make a *transistion* to a still lower energy state such as that represented by $E_c < 0$.

When the atoms have total energy E_c as drawn in this figure, they have only two turning points (labelled c in the figure). The classically permitted domain is now only the values of r in between these two points; everything less than the inner turning point or outside of the outer turning point corresponds to a kinetic energy that is less than 0 which is impossible. The classically forbidden regions for E_c are again shaded on the diagram. Atoms with this energy oscillate back and forth between these two turning points.

They oscillate back and forth very much like a *mass on a spring!* Note that this regions is labelled the ***quadratic region*** on the figure. This means that in this region, a quadratic function of $r - r_e$ (where r_e is the stable equilibrium at the minimum of $U(r)$ in this well) is a very good approximation to the actual potential energy. The potential energy of a mass on a spring aligned with r and with its equilibrium length moved so that it is r_e is just $\frac{1}{2}k(r-r_e)^2+U_0$, which can be *fit* to $U(r)$ in the quadratic region with a suitable choice of k and U_0.

Homework for Week 3

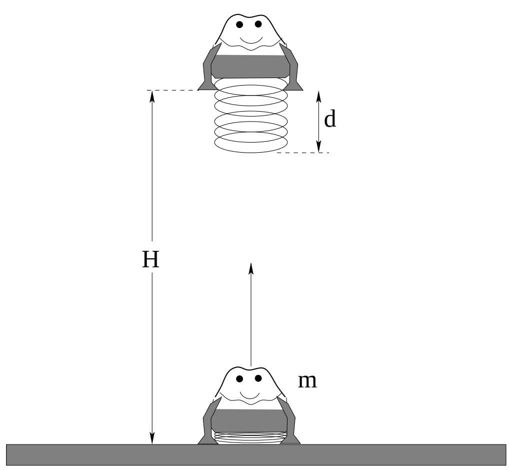
Problem 1.

Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

Problem 2.

Derive the Work-Kinetic Energy (WKE) theorem in one dimension from Newton's second law. You may use any approach used in class or given and discussed in this textbook (or any other), but do it ***yourself and without looking*** after studying.

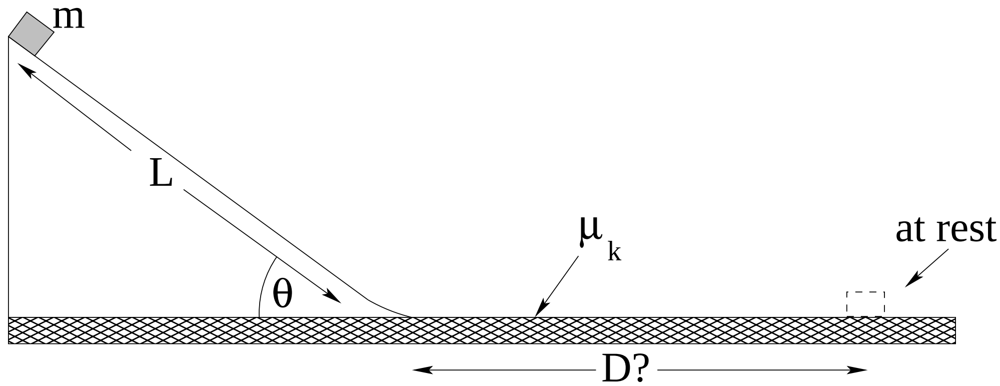
Problem 3.



A simple child's toy is a jumping frog made up of an approximately massless spring of uncompressed length d and spring constant k that propels a molded plastic "frog" of mass m. The frog is pressed down onto a table (compressing the spring by d) and at $t = 0$ the spring is released so that the frog leaps high into the air. Neglect drag and friction.

Use work and/or mechanical energy to determine how high the frog leaps.

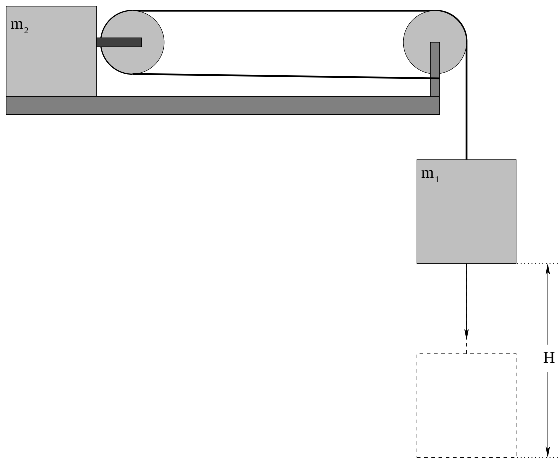
Problem 4.



A block of mass m, initially at rest, slides down a *smooth* (frictionless) incline of length L that makes an angle θ with the horizontal as shown. It then reaches a *rough* surface with a coefficient of kinetic friction μ_k.

Use the concepts of work and/or mechanical energy to find the distance D the block slides across the rough surface before it comes to rest. You will find that using the generalized non-conservative work-mechanical energy theorem is easiest, but you can succeed using work and mechanical energy conservation for two separate parts of the problem as well.

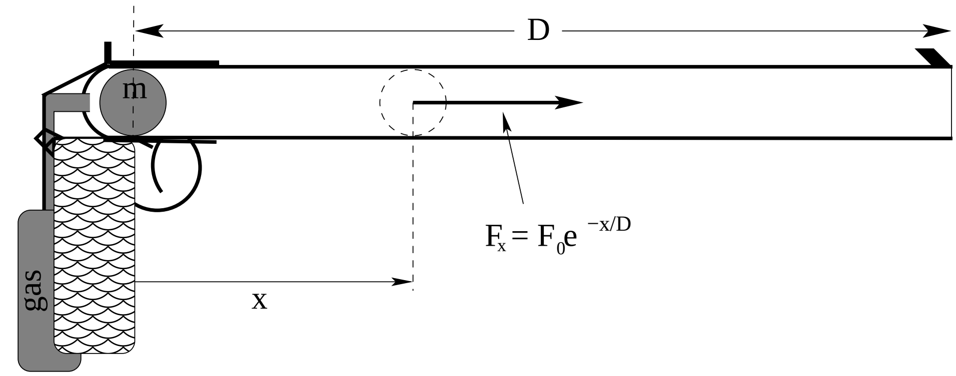
Problem 5.



A block of mass m_2 sits on a rough table. The coefficients of friction between the block and the table are μ_s and μ_k for static and kinetic friction respectively. A much larger mass m_1 (easily heavy enough to overcome static friction) is suspended from a massless, unstretchable, unbreakable rope that is looped around the two pulleys as shown and attached to the support of the rightmost pulley. At time $t = 0$ the system is released at rest.

Use work and/or mechanical energy (where the latter is *very easy* since the internal work done by the tension in the string *cancels*) to find the speed of both masses after the large mass m_1 has fallen a distance H. Note that you will still need to use the constraint between the coordinates that describe the two masses. Remember how hard you had to "work" to get this answer last week? When time isn't important, energy is better!

Problem 6.



A simple schematic for a paintball gun with a barrel of length D is shown above; when the trigger is pulled carbon dioxide gas under a fixed initial pressure is "suddenly" released into the approximately frictionless barrel behind the paintball (which has mass m). Suppose the rapidly expanding gas is cut off by a special valve so that it exerts a predictable ***total*** force on the ball of magnitude:

$$F_x = F_0 e^{-x/D}$$

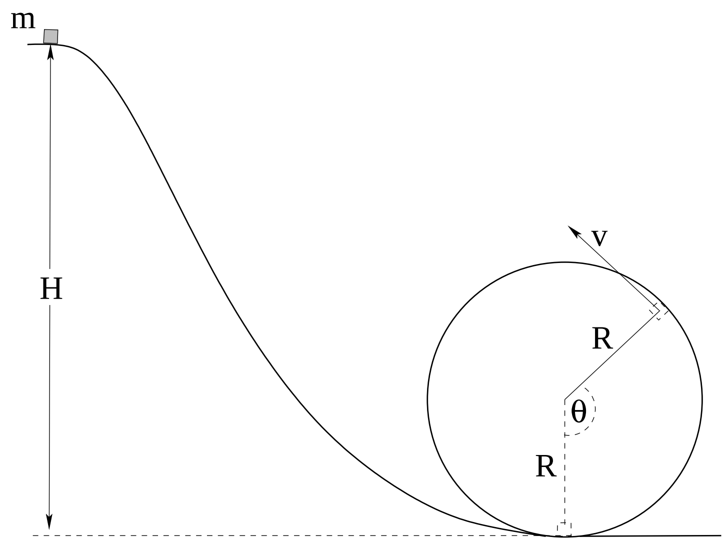
on the ball (where x is measured from the paintball's initial position as shown above) until the ball leaves the barrel.

a) Find the work done on the paintball by the force as the paintball is accelerated a total distance D down the barrel.

b) Use the work-kinetic-energy theorem to compute the kinetic energy of the paintball after it has been accelerated.

c) Find the speed with which the paintball emerges from the barrel after the trigger is pulled.

Problem 7.



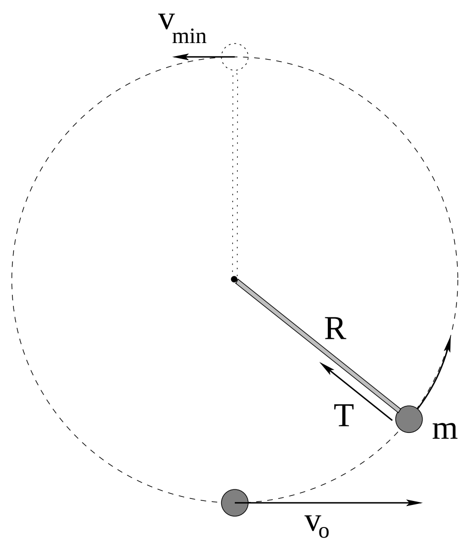
A block of mass M sits at rest at the top of a frictionless hill of height H leading to a circular frictionless loop-the-loop of radius R.

a) Find the minimum height $H_{\rm min}$ for which the block *barely* goes around the loop staying on the track at the top. (Hint: What is the condition on the normal force when it "barely" stays in contact with the track?).

b) Why was your answer to a) for $H_{\rm min}$ a dimensionless *number* times R and how this *kind* of result a good sign that your answer is probably right? You might want to reconsider your answer to a) if it is *not* of this form, by the way...

c) If the block is started at height $H_{\rm min}$, what is the normal force exerted by the track at the *bottom* of the loop where it is greatest?

Problem 8.



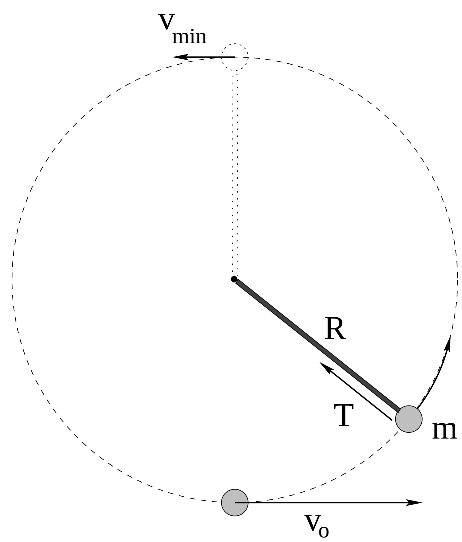
A ball of mass m is attached to a (massless, unstretchable) string and is suspended from a pivot. It is moving in a vertical circle of radius R such that it has speed v_0 at the bottom as shown. The ball is in a vacuum; neglect drag forces and friction in this problem. Near-Earth gravity acts down.

a) Find an expression for the force exerted on the ball by the string at the top of the loop as a function of m, g, R, and $v_{\rm top}$, assuming that the ball is still moving in a circle when it gets there.

b) Find the minimum speed $v_{\rm min}$ that the ball must have *at the top* to barely loop the loop (staying on the circular trajectory) with a precisely limp string with tension $T = 0$ at the top.

c) Determine the speed v_0 the ball must have *at the bottom* to arrive at the top with this minimum speed. You may use either work or potential energy for this part of the problem.

Problem 9.



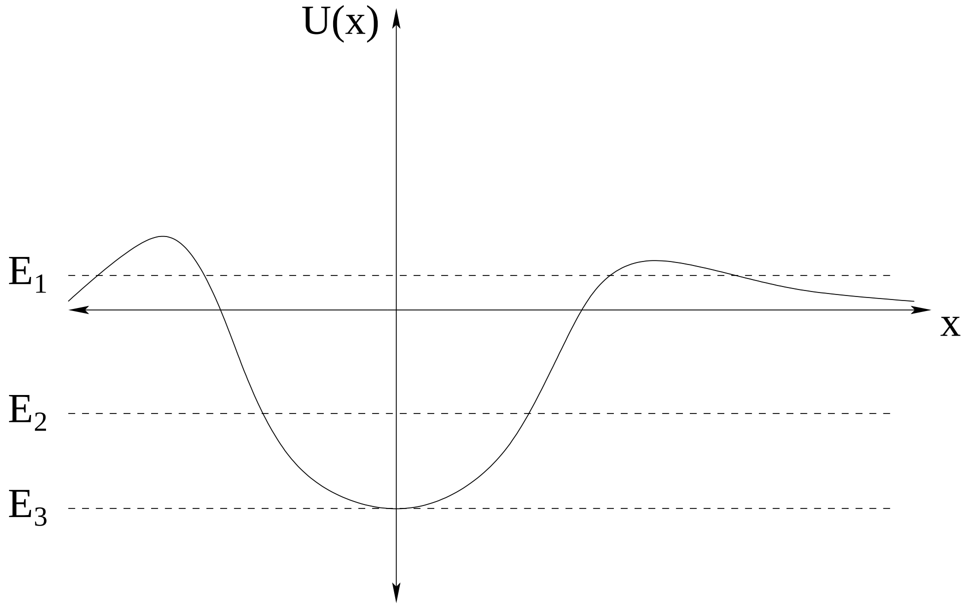
A ball of mass m is attached to a ***massless rod*** (note well) and is suspended from a frictionless pivot. It is moving in a vertical circle of radius R such that it has speed v_0 at the bottom as shown. The ball is in a vacuum; neglect drag forces and friction in this problem. Near-Earth gravity acts down.

a) Find an expression for the force exerted on the ball by the rod at the top of the loop as a function of m, g, R, and v_{top}, assuming that the ball is still moving in a circle when it gets there.

b) Find the minimum speed v_{min} that the ball must have *at the top* to barely loop the loop (staying on the circular trajectory). Note that this is easy, ***once you think about how the rod is different from a string!***

c) Determine the speed v_0 the ball must have *at the bottom* to arrive at the top with this minimum speed. You may use either work or potential energy for this part of the problem.

Problem 10.



a) On (a ***large*** copy of) the diagram above, place a small letter 'u' to mark points of unstable equilibrium.

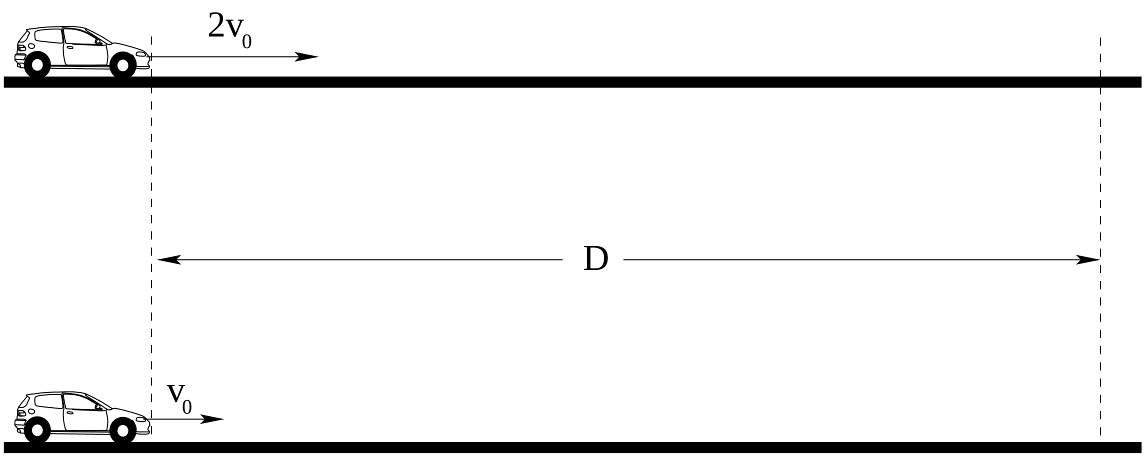
b) Place the letter 's' to mark points of stable equilibrium.

c) On the curve itself, place a few arrows in each distinct region indicating the ***direction*** of the force. Try to make the lengths of the arrows proportional in a *relative* way to the arrow you draw for the largest magnitude force.

d) For the three energies E_i shown, mark the *turning points* of motion with the letter 't_i' (e.g. t_1's for E_1, t_2's for E_2, t_3's for E_3.

e) For energy E_2, place $\overbrace{\qquad\qquad}^{\text{allowed}}$ to mark out the *classically allowed region* where the particle might be found. Place $\underbrace{\qquad\qquad}_{\text{forbidden}}$ to mark out the ***classically forbidden region*** where the particle can never be found.

Problem 11.



In the figure above we see two cars, one moving at a speed v_0 and an identical car moving at a speed $2v_0$. The cars are moving at a *constant* speed, so their motors are pushing them forward with a force that precisely cancels the drag force exerted by the air. This drag force is ***quadratic*** in their speed:

$$F_d = -cv^2$$

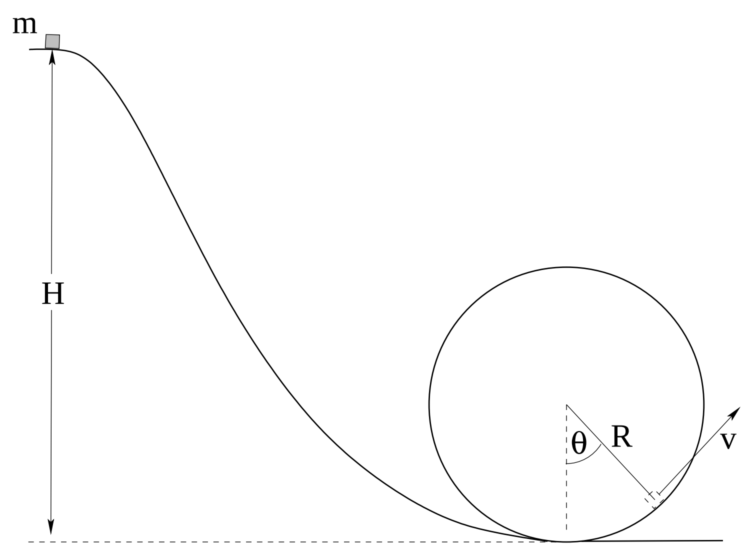
(in the opposite direction to their velocity) and we assume that this is the *only* force acting on the car in the direction of motion besides that provided by the motor itself, neglecting various other sources of friction or inefficiency.

a) Prove that the engine of the faster car has to be providing ***eight times as much power*** to maintain the higher constant speed than the slower car.

b) Prove that the faster car has to do ***four times as much work*** to travel a fixed distance D than the slower car. Note the difference between power and work!

Afterwards, you might want to think about (and possibly discuss with your physics-y or engineering friends): Cars typically *do* use more gasoline to drive the same distance at 100 kph (~ 62 mph) than they do at 50 kph, it isn't four times as much, or even twice as much. Why not? Things to consider include gears, engine efficiency, fuel wasted idling/stopped, friction, streamlining...

Advanced Problem 12.

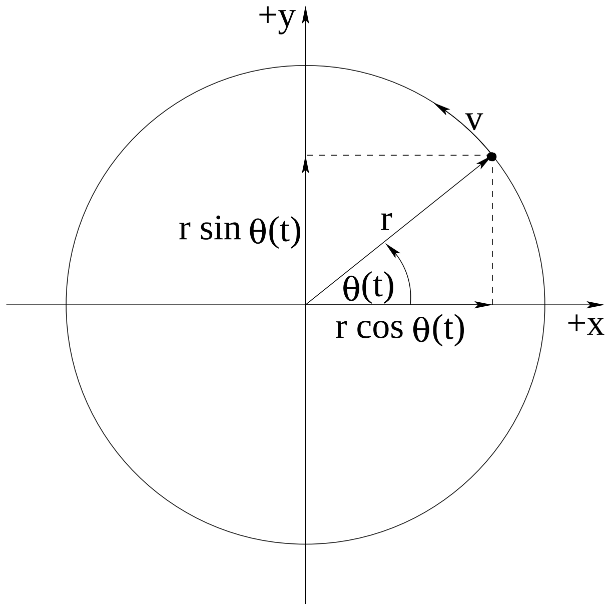


A block of mass M sits at the top of a frictionless hill of height H. It slides down and around a loop-the-loop of radius R, so that its position on the circle can be identified with the angle θ with respect to the vertical as shown

a) Find the magnitude of the normal force as a function of the angle θ.

b) From this, deduce an expression for the angle θ_0 at which the block will *leave* the track if the block is started at a height $H = 2R$.

Advanced Problem 13.



This is a guided exercise in ***calculus*** exploring the kinematics of circular motion and the relation between Cartesian and Plane Polar coordinates. It isn't as intuitive as the derivation given in the first two weeks, but it is much simpler and is formally correct.

In the figure above, note that:

$$\vec{\boldsymbol{r}} = r\cos\left(\theta(t)\right)\hat{\boldsymbol{x}} + r\sin\left(\theta(t)\right)\hat{\boldsymbol{y}}$$

where r is the radius of the circle and $\theta(t)$ is an *arbitrary* continuous function of time describing where a particle is on the circle at any given time. This is equivalent to:

$$\begin{aligned} x(t) &= r\cos(\theta(t)) \\ y(t) &= r\sin(\theta(t)) \end{aligned}$$

(going from (r,θ) plane polar coordinates to (x,y) cartesian coordinates and the corresponding:

$$\begin{aligned} r &= \sqrt{x(t)^2 + y(t)^2} \\ \theta(t) &= \tan^{-1}\left(\frac{y}{x}\right) \end{aligned}$$

You will find the following two definitions useful:

$$\begin{aligned} \Omega &= \frac{d\theta}{dt} \\ \alpha &= \frac{d\Omega}{dt} = \frac{d^2\theta}{dt^2} \end{aligned}$$

The first you should already be familiar with as the *angular velocity*, the second is the *angular acceleration*. Recall that the tangential speed $v_t = r\Omega$; similarly the tangential acceleration is $a_t = r\alpha$ as we shall see below.

Work through the following exercises. Starting with the observation that:

$$\vec{\boldsymbol{r}}(t)\cdot\vec{\boldsymbol{r}}(t) = \text{ constant}$$

is ***the*** assertion that the motion represented by $\vec{r}(t)$ is ***circular*** motion, whether or not it is at a uniform speed. Then:

a) Take the time derivative of this equation to prove that $\vec{v}(t) \perp \vec{r}(t)$.

b) Take one *more* time derivative of *this* equation and determine a (familiar) expression for the radial acceleration.

For the next two questions, use the *explicit* form for $\vec{r}(t)$ in cartesian coordinates.

c) Form the explicit dot product $\vec{v} \cdot \vec{r}$ in cartesian coordinates and show that it is *zero* (as shown more generally in a coordinate-free way above).

d) Take one more time derivative of this "zero" function in cartesian vector coordinates to show once again that:

$$\vec{a} = -\Omega^2\vec{r} + \frac{\alpha}{\Omega}\vec{v}$$

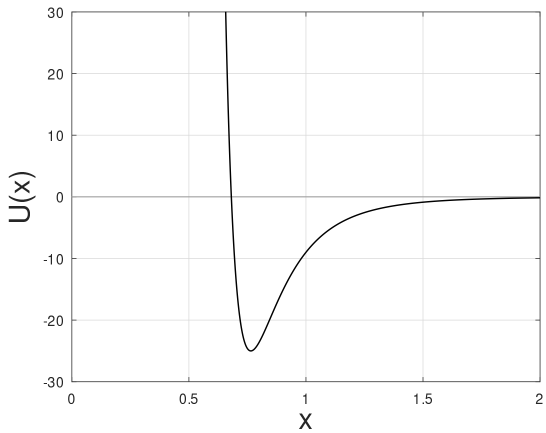
Since the direction of $\vec{v}$ is tangent to the circle of motion, we can identify these two terms as the results:

$$a_r = -\Omega^2 r = -\frac{v_t^2}{r}$$

(now derived in terms of its cartesian components) and

$$a_t = \alpha r.$$

Advanced Problem 14.



An object moves in the force produced by a potential energy function:

$$U(x) = \frac{1}{x^{12}} - \frac{10}{x^6}$$

This is a one-dimensional representation of an actual important physical interaction, in scaled (irrelevant) units, the one dimensional "12-6" ***Lennard-Jones*** potential energy function[109] . This function models dipole-induced dipole Van der Waals attraction at long range and a Pauli exclusion principle repulsion at short range between two atoms or molecules in a gas.

a) Find $x_{\rm eq}$, the *stable equilibrium distance of separation* of two e.g. noble gas atoms interacting with the Lennard-Jones potential. Two atoms separated by this distance at rest remain, stably, at rest!

b) Find $U(x_{\rm eq})$ the *binding energy* for two particles at rest separated by this distance. This is the mechanical energy one has to *add* to the system to cause this "molecule" to dissociate.

c) Find x_t, the *turning point distance* for $E = 0$. This is essentially the sum of the radii of the two atoms (in suitable coordinates – the scaled parameters used in this problem are not intended to be physical). Two molecules initially separated by a very long distance will be weakly attracted and fall towards each other, come momentarily to rest at this separation, and then push apart to infinity once again – *unless* they give off energy somehow and fall into a bound (negative energy) state!

d) Write an algebraic expression for $F_x(x)$. Note that it is strongly repulsive for $x < x_{\rm eq}$ and weakly attractive for all $x > x_{\rm eq}$.

[109]Wikipedia: http://www.wikipedia.org/wiki/Lennard-Jones potential. This is a good place for a wiki-romp for any students interested in physical chemistry. Note that this article links other important interaction functions and gives some explicit parameterizations. The Lennard-Jones potential turns out to be enormously useful in many contexts; hence its popularity.

Week 4: Systems of Particles, Momentum and Collisions

1.10: Summary

- **The center of mass** of a system of particles is given by:

$$\vec{\boldsymbol{x}}_{\rm cm} = \frac{\sum_i m_i \vec{\boldsymbol{x}}_i}{\sum_i m_i} = \frac{1}{M_{\rm tot}} \sum_i m_i \vec{\boldsymbol{x}}_i$$

  One can differentiate this expression once or twice with respect to time to get the two corollary expressions:

$$\vec{\boldsymbol{v}}_{\rm cm} = \frac{\sum_i m_i \vec{\boldsymbol{v}}_i}{\sum_i m_i} = \frac{1}{M_{\rm tot}} \sum_i m_i \vec{\boldsymbol{v}}_i$$

  and

$$\vec{\boldsymbol{a}}_{\rm cm} = \frac{\sum_i m_i \vec{\boldsymbol{a}}_i}{\sum_i m_i} = \frac{1}{M_{\rm tot}} \sum_i m_i \vec{\boldsymbol{a}}_i$$

  All three expressions may be summed up in the useful forms:

$$\begin{aligned} M_{\rm tot}\vec{\boldsymbol{x}}_{\rm cm} &= \sum_i m_i \vec{\boldsymbol{x}}_i \\ M_{\rm tot}\vec{\boldsymbol{v}}_{\rm cm} &= \sum_i m_i \vec{\boldsymbol{v}}_i \\ M_{\rm tot}\vec{\boldsymbol{a}}_{\rm cm} &= \sum_i m_i \vec{\boldsymbol{a}}_i \end{aligned}$$

  The center of mass coordinates are truly *weighted averages* of the coordinates – weighted with the *actual weights* of the particles[110].

- **The mass density** of a solid object in one, two, or three dimensions is traditionally written in physics as:

$$\begin{aligned} \mu &= \lim_{\Delta x \to 0} \frac{\Delta m}{\Delta x} = \frac{dm}{dx} \\ \sigma &= \lim_{\Delta x \to 0} \frac{\Delta m}{\Delta A} = \frac{dm}{dA} \\ \rho &= \lim_{\Delta x \to 0} \frac{\Delta m}{\Delta V} = \frac{dm}{dV} \end{aligned}$$

[110] Near the Earth's surface where the weight only depends on the mass, of course. Really they are weighted with the mass.

In each of these expressions, Δm is the mass in a small "chunk" of the material, one of length Δx, area ΔA, or volume ΔV. The mass distribution of an object is in general ***a complicated function of the coordinates***[111]. However we will usually work only with *very simple* mass distributions that we can easily integrate/sum over in this class. When doing so we are likely to use these definitions *backwards*:

$$\begin{aligned} dm &= \mu\, dx \qquad 1 \text{ dimension} \\ dm &= \sigma\, dA \qquad 2 \text{ dimensions} \\ dm &= \rho\, dV \qquad 3 \text{ dimensions} \end{aligned}$$

Use the following ritual incantation (which will be useful to you repeatedly for both semesters of this course!) when working with mass (or later, charge) density distributions:

> **The mass of the chunk is the mass per unit (length, area, volume) times the (length, area, volume) of the chunk!**

- **The Center of Mass** of a solid object (continuous mass distribution) is given by:

$$\vec{x}_{\rm cm} = \frac{\int \vec{x}\, dm}{\int dm} = \frac{\int \vec{x}\, \rho(\vec{x})\, dV}{\int \rho(\vec{x})\, dV} = \frac{1}{M_{\rm tot}} \int \vec{x}\, \rho(\vec{x})\, dV$$

This can be evaluated one component at a time, e.g.:

$$x_{\rm cm} = \frac{\int x\, dm}{\int dm} = \frac{\int x\, \rho(\vec{x})\, dV}{\int \rho(\vec{x})\, dV} = \frac{1}{M_{\rm tot}} \int x\, \rho(\vec{x})\, dV$$

(and similarly for $y_{\rm cm}$ and $z_{\rm cm}$).

It also can be written (componentwise) for mass distributions in one and two dimensions:

$$x_{\rm cm} = \frac{\int x\, dm}{\int dm} = \frac{\int x\, \mu\, dx}{\int \mu\, dx} = \frac{1}{M_{\rm tot}} \int x\, \mu\, dx$$

(in one dimension) or

$$x_{\rm cm} = \frac{\int x\, dm}{\int dm} = \frac{\int x\, \sigma\, dA}{\int \sigma\, dA} = \frac{1}{M_{\rm tot}} \int x\, \sigma\, dA$$

and

$$y_{\rm cm} = \frac{\int y\, dm}{\int dm} = \frac{\int y\, \sigma\, dA}{\int \sigma\, dA} = \frac{1}{M_{\rm tot}} \int y\, \sigma\, dA$$

(in two dimensions).

- **The Momentum** of a particle is **defined** to be:

$$\vec{p} = m\vec{v}$$

The momentum of a **system of particles** is the sum of the momenta of the individual particles:

$$\vec{p}_{\rm tot} = \sum_i m_i \vec{v}_i = \sum m_i \vec{v}_{\rm cm} = M_{\rm tot} \vec{v}_{\rm cm}$$

where the last expression follows from the expression for the velocity of the center of mass above.

[111] Think about how mass is distributed in the human body! Or, for that matter, think about the Universe itself, which can be thought of at least partially as a great big mass density distribution $\rho(\vec{x})$...

- **The Kinetic Energy in Terms of the Momentum** of a particle is easily written as:

$$K = \frac{1}{2}mv^2 = \frac{1}{2}mv^2\left(\frac{m}{m}\right) = \frac{(mv)^2}{2m} = \frac{p^2}{2m}$$

  or (for a system of particles):

$$K_{\rm tot} = \sum_i \frac{1}{2}m_i v_i^2 = \sum_i \frac{p_i^2}{2m_i}$$

  These forms are ***very useful in collision problems*** where momentum is known and conserved; they will often save you a step or two in the algebra if you express kinetic energies in terms of momenta from the beginning.

- **Newton's Second Law** for a single particle can be expressed (and was so expressed, originally, by Newton) as:

$$\vec{\boldsymbol{F}}_{\rm tot} = \frac{d\vec{\boldsymbol{p}}}{dt}$$

  where $\vec{\boldsymbol{F}}_{\rm tot}$ is the total force acting on the particle.

  For a system of particles one can sum this:

$$\vec{\boldsymbol{F}}_{\rm tot} = \sum_i \vec{\boldsymbol{F}}_i = \sum_i \frac{d\vec{\boldsymbol{p}}_i}{dt} = \frac{d\sum_i \vec{\boldsymbol{p}}_i}{dt} = \frac{d\vec{\boldsymbol{p}}_{\rm tot}}{dt}$$

  In this expression the *internal forces directed along the lines* between particles of the system cancel (due to Newton's Third Law) and:

$$\vec{\boldsymbol{F}}_{\rm tot} = \sum_i \vec{\boldsymbol{F}}_i^{\rm ext} = \frac{d\vec{\boldsymbol{p}}_{\rm tot}}{dt}$$

  where the total force in this expression is the sum of only the total *external* forces acting on the various particles of the system.

- **The Law of Conservation of Momentum** states (following the previous result) that:

  > **If and only if the total external force acting on a system of particles vanishes, then the total momentum of that system is a constant vector.**

  or (in equationspeak):

$$\text{If and only if } \vec{\boldsymbol{F}}_{\rm tot} = 0 \text{ then } \vec{\boldsymbol{p}}_{\rm tot} = \vec{\boldsymbol{p}}_i = \vec{\boldsymbol{p}}_f, \text{ a constant vector}$$

  where $\vec{\boldsymbol{p}}_i$ and $\vec{\boldsymbol{p}}_f$ are the initial and final momenta across some intervening process or time interval where no external forces acted. Momentum conservation is especially useful in ***collision problems*** because the collision force is internal and hence does not change the total momentum.

- **The Center of Mass Reference Frame** is a convenient frame for solving collision problems. It is the frame whose origin lies at the center of mass and that moves at the constant velocity (relative to "the lab frame") of the center of mass. That is, it is the frame wherein:

$$\vec{\boldsymbol{x}}_i' = \vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_{\rm cm} = \vec{\boldsymbol{x}}_i - \vec{\boldsymbol{v}}_{\rm cm}t$$

and (differentiating once):

$$\vec{v}_i' = \vec{v}_i - \vec{v}_{\rm cm}$$

In this frame,

$$\vec{p}_{\rm tot}' = \sum_i m_i \vec{v}_i' = \sum_i m_i \vec{v}_i - \sum_i m_i \vec{v}_{\rm cm} = \vec{p}_{\rm tot} - \vec{p}_{\rm tot} = 0$$

which is why it is so very useful. The total momentum is the constant value *0* in the center of mass frame of a system of particles with no external forces acting on it!

- **The Impulse** of a collision is defined to be the ***total momentum transferred*** during the collision, where a collision is an event where a very large force is exerted over a very short time interval Δt. Recalling that $\vec{F} = d\vec{p}/dt$, it's magnitude is:

  $$I = |\Delta \vec{p}| = \left| \int_0^{\Delta t} \vec{F}\ dt \right| = |\vec{F}_{\rm avg}|\Delta t$$

  and it usually acts along the line of the collision. Note that this the impulse is directly related to the ***average force*** exerted by a collision that lasts a very short time Δt:

  $$\vec{F}_{\rm avg} = \frac{1}{\Delta t} \int_0^{\Delta t} \vec{F}(t)\ dt$$

- **An Elastic Collision** is by definition a collision in which ***both*** the momentum ***and*** the total kinetic energy of the particles is conserved across the collision. That is:

  $$\begin{aligned} \vec{p}_i &= \vec{p}_f \\ K_i &= K_f \end{aligned}$$

  This is actually *four independent conservation equations* (three components of momentum and kinetic energy).

  In general we will be given six "initial values" for a three-dimensional collision – the three components of the initial velocity for each particle. Our goal is to find the six final values – the three components of the final velocity of each particle. However, *we don't have enough simultaneous equations to accomplish this* and therefore have to be given two more pieces of information in order to solve a general elastic collision problem in three dimensions.

  In two dimensional collisions we are a bit better off – we have three conservation equations (two momenta, one energy) and four unknowns (four components of the final velocity) and can solve the collision if we know one more number, say the *angle* at which one of the particles emerges or the *impact parameter* of the collision[112] , but it is still pretty difficult.

  In one dimension we have two conservation equations – one momentum, one energy, and two unknowns (the two final velocities) and we can (almost) uniquely solve for the final velocities given the initial ones. In this latter case only, when the initial state of the two particles is given by m_1, v_{1i}, m_2, v_{2i} then the final state is given by:

  $$\begin{aligned} v_{1f} &= -v_{1i} + 2v_{\rm cm} \\ v_{2f} &= -v_{2i} + 2v_{\rm cm} \end{aligned}$$

[112] Wikipedia: http://www.wikipedia.org/wiki/impact parameter.

- **An Inelastic Collision** is by definition not an elastic collision, that is, a collision where ***kinetic energy is not conserved***. Note well that the term "elastic" therefore refers to *conservation of energy* which may or may not be present in a collision, but that ***MOMENTUM IS ALWAYS CONSERVED IN A COLLISION*** in the impact approximation, which we will universally make in this course.

  A ***fully inelastic collision*** is one where the two particles collide and ***stick together*** to move as one after the collision. In three dimensions we therefore have three conserved quantities (the components of the momentum) and three unknown quantities (the three components of the final velocity and therefore ***fully inelastic collisions are trivial to solve!*** The solution is simply to find:

$$\vec{\boldsymbol{P}}_{\rm tot} = \vec{\boldsymbol{P}}_i = m_1\vec{\boldsymbol{v}}_{1,i} + m_2\vec{\boldsymbol{v}}_{2,i}$$

  and set it equal to $\vec{\boldsymbol{P}}_f$:

$$m_1\vec{\boldsymbol{v}}_{1,i} + m_2\vec{\boldsymbol{v}}_{2,i} = (m_1+m_2)\vec{\boldsymbol{v}}_f = (m_1+m_2)\vec{\boldsymbol{v}}_{\rm cm}$$

  or

$$\vec{\boldsymbol{v}}_f = \vec{\boldsymbol{v}}_{\rm cm} = \frac{m_1\vec{\boldsymbol{v}}_{1,i} + m_2\vec{\boldsymbol{v}}_{2,i}}{m_1+m_2} = \frac{\vec{\boldsymbol{P}}_{\rm tot}}{M_{\rm tot}}$$

  The final velocity of the stuck together masses is the (constant) velocity of the center of mass of the system, which makes complete sense.

  Kinetic energy is always lost in an inelastic collision, and one can always evaluate it from:

$$\Delta K = K_f - K_i = \frac{P_{\rm tot}^2}{2M_{\rm tot}} - \left(\frac{p_{1,i}^2}{2m_1} + \frac{p_{2,i}^2}{2m_2}\right)$$

  In a *partially* inelastic collision, the particles collide but don't quite stick together. One has three (momentum) conservation equations and needs six final velocities, so one in general must be given three pieces of information in order to solve a partially inelastic collision in three dimensions. Even in one dimension one has only one equation and two unknowns and hence one needs at least one additional piece of independent information to solve a problem.

- The Kinetic Energy of a System of Particles can in general be written as:

$$K_{\rm tot} = \left(\sum_i K_i'\right) + K_{\rm cm} = K_{\rm tot}' + K_{\rm cm}$$

  which one should read as "The total kinetic energy of a system in the lab frame equals its total kinetic energy ***in*** the (primed) center of mass frame plus the kinetic energy ***of*** the center of mass frame treated as a 'particle' in the lab frame."

  The kinetic energy *of* the center of mass frame in the lab is thus just:

$$K_{\rm cm} = \frac{1}{2}M_{\rm tot}v_{\rm cm}^2 = \frac{P_{\rm tot}^2}{2M_{\rm tot}}$$

  which we recognize as "the kinetic energy of a baseball treated as a particle with its total mass located at its center of mass" (for example) even though the baseball is *really* made up of many, many small particles that generally have kinetic energy of their own relative to the center of mass.

This theorem will prove very useful to us when we consider rotation, but it also means that the total kinetic energy of a macroscopic object (such as a baseball) made up of many microscopic parts is the *sum* of its macroscopic kinetic energy – its kinetic energy where we treat it as a "particle" located at its center of mass – and its ***internal microscopic kinetic energy***. The latter is essentially related to *enthalpy, heat and temperature*. Inelastic collisions that "lose kinetic energy" of their macroscopic constituents (e.g. cars) gain it in the *increase in temperature* of the objects after the collision that results from the greater microscopic kinetic energy of the particles that make them up in the center of mass (object) frame.

4.1: Systems of Particles

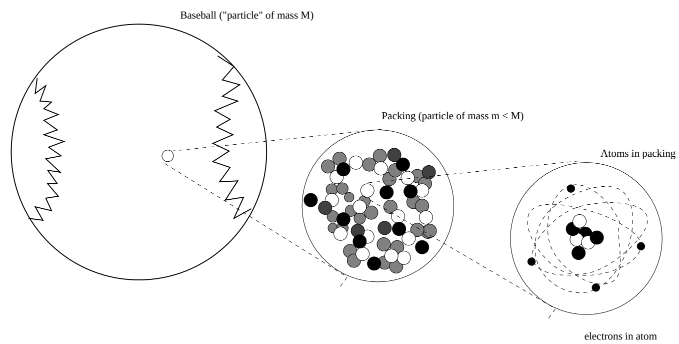


Figure 4.1: An object such as a baseball is not really a particle. It is made of *many, many* particles – even the atoms it is made of are made of many particles *each*. Yet it *behaves* like a particle as far as Newton's Laws are concerned. Now we find out why.

The world of one particle as we've learned it so far is fairly simple. Something pushes on it, and it accelerates, its velocity changing over time. Stop pushing, it coasts or remains still with its velocity constant. Or from another (time independent) point of view: Do work on it and it speeds up. Do negative work on it and it slows down. Increase or decrease its potential energy; decrease or increase its kinetic energy.

However, the world of *many* particles is not so simple. For one thing, ***every push works two ways*** – all forces act symmetrically between objects – ***no object experiences a force all by itself***. For another, real objects are not particles – they are made up of lots of "particles" themselves. Finally, even if we ignore the internal constituents of an object, we seem to inhabit a universe with lots of *macroscopic* objects. If we restrict ourselves to objects the size of *stars* there are well over a ***hundred billion stars in our Milky Way galaxy*** (which is fairly average as far as size and structure are concerned) and there are well over a ***hundred billion galaxies*** visible to the Hubble, meaning that there are ***at least 10^{20} stars*** visible to our instruments. One can get quite bored writing out the zeros in a number like that even before we consider just how many electrons and quarks each star (on average) is made up of!

Somehow we know intuitively that the details of the motion of every electron and quark in a baseball, or a star, are irrelevant to the motion and behavior of the baseball/star as a whole, treated as a "particle" itself. Clearly, we need to deduce ways of taking a collection of particles and determining its collective behavior. Ideally, this process should be one we can iterate, so that we can treat collections of collections – a box of baseballs, under the right circumstances (falling out of an airplane, for example) might also be expected to behave within reason like a single object independent of the motion of the baseballs inside, or the motion of the atoms in the baseballs, or the motion of the electrons and quarks in the atoms.

We will obtain this collective behavior by **averaging**, or *summing over* (at successively larger scales) the physics that we know applies at the smallest scale to things that *really are* particles and discover to our surprise that it applies equally well to collections of those particles, subject to a few new definitions and rules.

4.1.1: Newton's Laws for a System of Particles – Center of Mass

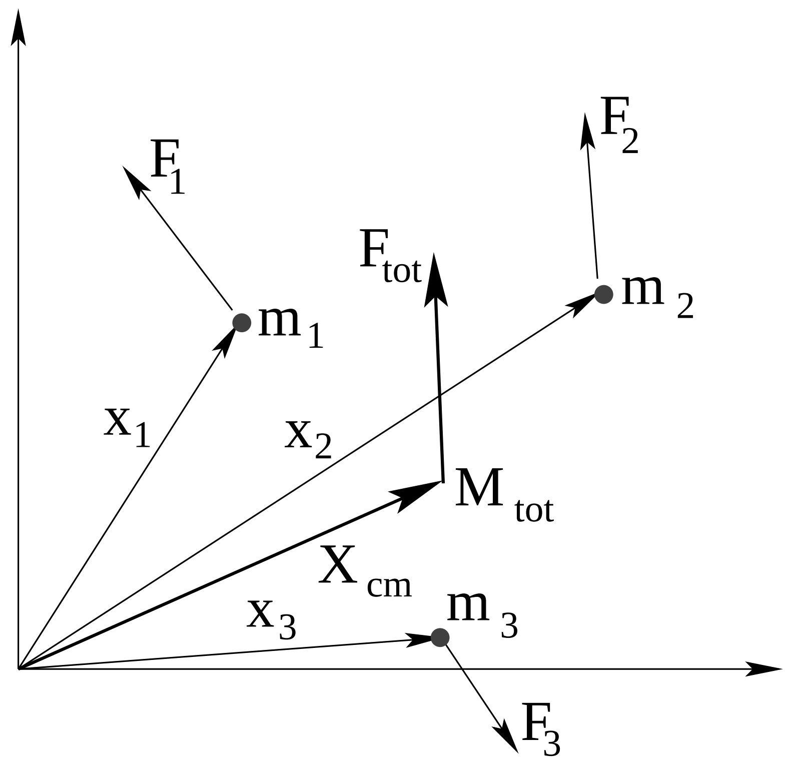


Figure 4.2: A system of $N = 3$ particles is shown above, with various forces $\vec{F}_i$ acting on the masses (which therefore each their own accelerations $\vec{a}_i$). From this, we construct a *weighted average* acceleration of the system, in such a way that Newton's Second Law is satisfied for the *total* mass.

Suppose we have a system of N particles, each of which is experiencing a force. Some (part) of those forces are "external" – they come from outside of the system. Some (part) of them may be "internal" – equal and opposite force pairs between particles that help hold the system together (solid) or allow its component parts to interact (liquid or gas).

We would like the total force to act on the total mass of this system as if it were a "particle". That is, we would like for:

$$\vec{F}_{\rm tot} = M_{\rm tot}\vec{A} \tag{4.1}$$

where $\vec{A}$ is the "acceleration of the system". This is easily accomplished.

Newton's Second Law for a system of particles is written as:

$$\vec{F}_{\rm tot} = \sum_i \vec{F}_i = \sum_i m_i \frac{d^2\vec{x}_i}{dt^2} \tag{4.2}$$

We now perform the following Algebra Magic:

$$\begin{aligned}
\vec{F}_{\rm tot} &= \sum_i \vec{F}_i && (4.3)\\
&= \sum_i m_i \frac{d^2\vec{x}_i}{dt^2} && (4.4)\\
&= \left(\sum_i m_i\right)\frac{d^2\vec{X}}{dt^2} && (4.5)\\
&= M_{\rm tot}\frac{d^2\vec{X}}{dt^2} = M_{\rm tot}\vec{A} && (4.6)
\end{aligned}$$

Note well the introduction of a new coordinate, $\vec{X}$. This introduction isn't "algebra", it is a *definition*. Let's isolate it so that we can see it better:

$$\sum_i m_i \frac{d^2\vec{x}_i}{dt^2} = M_{\rm tot}\frac{d^2\vec{X}}{dt^2} \tag{4.7}$$

Basically, *if* we define an $\vec{X}$ such that this relation is true *then* Newton's second law is recovered for the entire system of particles "located at $\vec{X}$" as if that location were indeed a particle of mass $M_{\rm tot}$ itself.

We can rearrange this a bit as:

$$\frac{d\vec{V}}{dt} = \frac{d^2\vec{X}}{dt^2} = \frac{1}{M_{\rm tot}}\sum_i m_i \frac{d^2\vec{x}_i}{dt^2} = \frac{1}{M_{\rm tot}}\sum_i m_i \frac{d\vec{v}_i}{dt} \tag{4.8}$$

and can *integrate twice* on both sides (as usual, but we only do the integrals formally). The first integral is:

$$\frac{d\vec{X}}{dt} = \vec{V} = \frac{1}{M_{\rm tot}}\sum_i m_i\vec{v}_i + \vec{V}_0 = \frac{1}{M_{\rm tot}}\sum_i m_i\frac{d\vec{x}_i}{dt} + \vec{V}_0 \tag{4.9}$$

and the second is:

$$\vec{X} = \frac{1}{M_{\rm tot}}\sum_i m_i\vec{x}_i + \vec{V}_0 t + \vec{X}_0 \tag{4.10}$$

Note that this equation is exact, but we have had to introduce *two constants of integration* that are completely arbitrary: $\vec{V}_0$ and $\vec{X}_0$.

These constants represent the exact same freedom that we have with our *inertial frame of reference* – we can put the origin of coordinates anywhere we like, and we will get the same equations of motion even if we put it somewhere and describe everything in a uniformly moving frame. We should have *expected* this sort of freedom in our definition of a coordinate that describes "the system" because we have precisely the same freedom in our choice of *coordinate system* in terms of which to describe it.

In many problems, however, we don't want to use this freedom. Rather, we want the *simplest* description of the system itself, and push all of the freedom concerning constants of motion over to the coordinate choice itself (where it arguably "belongs"). We therefore select just *one* (the simplest one) of the infinity of possibly consistent rules represented in our definition above that would preserve Newton's Second Law and call it by a special name: ***The Center of Mass!***

We define the position of the center of mass to be:

$$M\vec{\boldsymbol{X}}_{\rm cm} = \sum_i m_i \vec{\boldsymbol{x}}_i \tag{4.11}$$

or:

$$\vec{\boldsymbol{X}}_{\rm cm} = \frac{1}{M}\sum_i m_i \vec{\boldsymbol{x}}_i \tag{4.12}$$

(with $M = \sum_i m_i$). If we consider the "location" of the system of particles to be the center of mass, then Newton's Second Law will be satisfied for the system as if it were a particle, and the location in question will be exactly what we intuitively expect: the "middle" of the (collective) object or system, weighted by its distribution of mass.

Not all systems we treat will appear to be made up of point particles. Most solid objects or fluids appear to be made up of a *continuum* of mass, a *mass distribution*. In this case we need to do the sum by means of *integration*, and our definition becomes:

$$M\vec{\boldsymbol{X}}_{\rm cm} = \int \vec{\boldsymbol{x}} dm \tag{4.13}$$

or

$$\vec{\boldsymbol{X}}_{\rm cm} = \frac{1}{M}\int \vec{\boldsymbol{x}} dm \tag{4.14}$$

(with $M = \int dm$). The latter form comes from treating every little differential chunk of a solid object like a "particle", and adding them all up. Integration, recall, is just a way of adding them up.

Of course this leaves us with the recursive problem of the fact that "solid" objects are *really* made out of lots of point-like elementary particles and their fields. It is worth very briefly presenting the standard "coarse-graining" argument that permits us to treat solids and fluids like a continuum of smoothly distributed mass – and the limitations of that argument.

Example 4.1.1: Center of Mass of a Few Discrete Particles

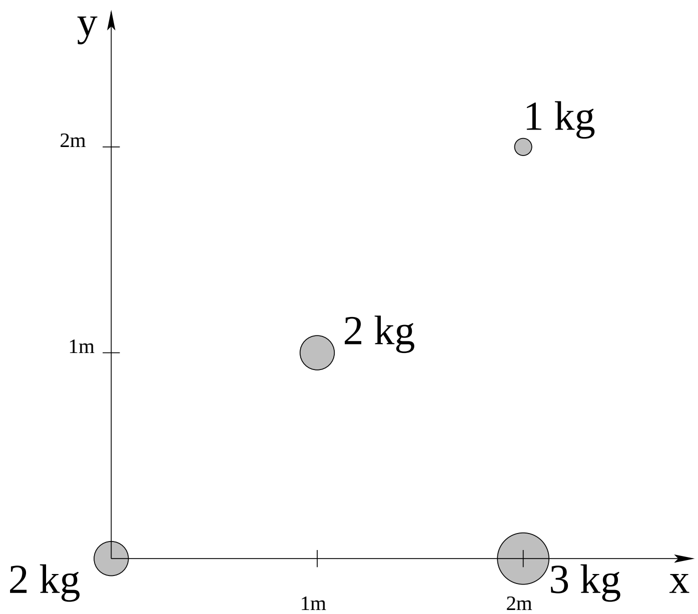


Figure 4.3: A system of four massive particles.

In figure 4.3 above, a few discrete particles with masses given are located at the positions indicated. We would like to find the center of mass of this system of particles. We do this by arithmetically evaluating the algebraic expressions for the x and y components of the center of mass separately:

$$x_{\rm cm} = \frac{1}{M_{\rm tot}} \sum_i m_i x_i = \frac{1}{8}(2*0+2*1+3*2+1*2) = 1.25 \text{ m} \tag{4.15}$$

$$y_{\rm cm} = \frac{1}{M_{\rm tot}} \sum_i m_i y_i = \frac{1}{8}(2*0+3*0+2*1+1*2) = 0.5 \text{ m} \tag{4.16}$$

Hence the center of mass of this system is located at $\vec{x}_{\rm cm} = 1.25\hat{x} + 0.5\hat{y}$.

4.1.2: Coarse Graining: Continuous Mass Distributions

Suppose we wish to find the center of mass of a small cube of some uniform material – such as gold, why not? We know that *really* gold is made up of gold atoms, and that gold atoms are make up of (elementary) electrons, quarks, and various massless field particles that bind the massive particles together. In a cube of gold with a mass of 197 grams, there are roughly 6×10^{23} atoms, each with 79 electrons and 591 quarks for a total of 670 elementary particles per atom. This is then about 4×10^{26} elementary particles in a cube just over 2 cm per side.

If we tried to actually *use the sum form* of the definition of center of mass to evaluate it's location, and ran the computation on a computer capable of performing one trillion floating point operations per second, it would take several hundred trillion seconds (say ten million years) and – unless we knew the exacly location of every quark – would *still* be approximate, no better than a guess.

We do far better by *averaging*. Suppose we take a small chunk of the cube of gold – one with cube edges 1 millimeter long, for example. This still has an enormous number of elementary particles in it – so many that if we shift the boundaries of the chunk a tiny bit *many* particles – many *whole atoms* are moved in or out of the chunk. Clearly we are justified in talking about the "average number of atoms" or "average amount of mass of gold" in a tiny cube like this.

A millimeter is still absurdly large on an atomic scale. We could make the cube 1 *micron* (1×10^{-6} meter, a thousandth of a millimeter) and because atoms have a "generic" size around one Angstrom – 1×10^{-10} meters – we would expect it to contain around $(10^{-6}/10^{-10})^3 = 10^{12}$ atoms. Roughly a trillion atoms in a cube too small to see with the naked eye (and each atom still has almost 700 elementary particles, recall). We could go down at least 1-2 *more* orders of magnitude in size and still have millions of particles in our chunk!

A chunk 10 nanometers to the side is fairly accurately located in space on a scale of meters. It has enough elementary particles in it that we can meaningfully speak of its "average mass" and use this to define the *mass density* at the point of location of the chunk – the mass per unit volume at that point in space – with at least 5 or 6 significant figures (one part in a million accuracy). In most real-number computations we might undertake in the kind of physics learned in this class, we wouldn't pay attention to more than 3 or 4 significant figures, so this is plenty.

The point is that this chunk is now small enough to be considered *differentially small* for the purposes of doing *calculus*. This is called *coarse graining* – treating chunks big on an atomic or molecular scale but small on a macroscopic scale. To complete the argument, in physics we would generally consider a small chunk of matter in a solid or fluid that we wish to treat as a smooth distribution of mass, and write at first:

$$\Delta m = \rho \Delta V \tag{4.17}$$

while reciting the following magical formula to ourselves:

> **The mass of the chunk is the mass per unit volume ρ times the volume of the chunk.**

We would then think to ourselves: "Gee, ρ is *almost* a uniform function of location for chunks that are small enough to be considered a differential as far as doing sums using integrals are concerned. I'll just coarse grain this and use integration to evaluation all sums." Thus:

$$dm = \rho \, dV \tag{4.18}$$

We do this all of the time, in this course. This semester we do it repeatedly for mass distributions, and sometimes (e.g. when treating planets) will coarse grain on a much larger scale to form the "average" density on a planetary scale. On a planetary scale, barring chunks of neutronium or the occasional black hole, a cubic kilometer "chunk" is still "small" enough to be considered differentially small – we usually won't need to integrate over every single distinct pebble or clod of dirt on a much smaller scale. Next semester we will do it repeatedly for electrical charge, as after all all of those gold atoms are made up of *charged* particles so there are just as many charges to consider as there are elementary particles. Our models for electrostatic fields of continuous charge and electrical currents in wires will all rely on this sort of coarse graining.

Before we move on, we should say a word or two about two other common distributions of mass. If we want to find e.g. the center of mass of a flat piece of paper cut out into (say) the shape of a triangle, we *could* treat it as a "volume" of paper and integrate over its thickness. However, it is probably a pretty good bet from *symmetry* that unless the paper is very inhomogeneous across its thickness, the center of mass in the flat plane is in the middle of the "slab" of paper, and the paper is already so thin that we don't pay much attention to its thickness as a general rule. In this case we basically integrate out the thickness in our minds (by multiplying ρ by the paper thickness t) and get:

$$\Delta m = \rho \Delta V = \rho t \Delta A = \sigma \Delta A \tag{4.19}$$

where $\sigma = \rho t$ is the (average) *mass per unit area* of a chunk of paper with *area* ΔA. We say our (slightly modified) magic ritual and poof! We have:

$$dm = \sigma \, dA \tag{4.20}$$

for two dimensional *areal distributions* of mass.

Similarly, we often will want to find the center of mass of things like wires bent into curves, things that are *long and thin*. By now I shouldn't have to explain the following reasoning:

$$\Delta m = \rho \Delta V = \rho A \Delta x = \mu \Delta x \tag{4.21}$$

where A is the (small!) cross section of the solid wire and $\mu = \rho A$ is the *mass per unit length* of the chunk of wire, magic spell, cloud of smoke, and when the smoke clears we are left with:

$$dm = \mu \, dx \tag{4.22}$$

In all of these cases, note well, ρ, σ, μ can be *functions of the coordinates!* They are not necessarily *constant*, they simply describe the (average) mass per unit volume at the point in our object or system in question, subject to the coarse-graining limits. Those limits are pretty sensible ones – if we are trying to solve problems on a length scale of angstroms, we *cannot use these averages* because the laws of large numbers won't apply. Or rather, we can and do still use these *kinds* of averages in quantum theory (because even on the scale of a *single atom* doing all of the discrete computations proves to be a problem) but then we do so knowing up front that they are approximations and that our answer will be "wrong".

In order to use the idea of center of mass (CM) in a problem, we need to be able to evaluate it. For a system of discrete particles, the sum definition is all that there is – you brute-force your way through the sum (decomposing vectors into suitable coordinates and adding them up).

For a solid object that is symmetric, the CM is "in the middle". But where's that? To precisely find out, we have to be able to use the integral definition of the CM:

$$M\vec{\boldsymbol{X}}_{\rm cm} = \int \vec{\boldsymbol{x}} dm \tag{4.23}$$

(with $M = \int dm$, and $dm = \rho dV$ or $dm = \sigma dA$ or $dm = \mu dl$ as discussed above).

Let's try a few examples:

Example 4.1.2: Center of Mass of a Continuous Rod

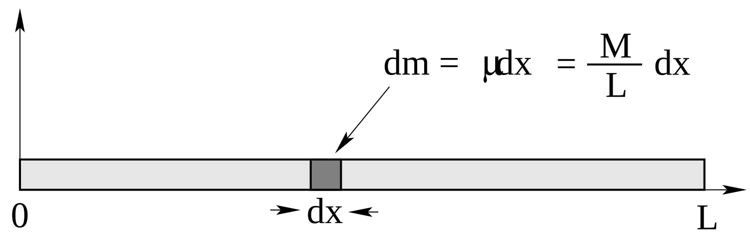


Figure 4.4

Let us evaluate the center of mass of a ***thin*** continuous rod of length L and total mass M, to make sure it is in the middle:

$$M x_{\mathrm{cm}} = \int \vec{x} dm = \int_0^L x(\mu\, dx) = \frac{1}{2}\mu L^2 \tag{4.24}$$

where

$$M = \int dm = \int_0^L \mu dx = \mu L \tag{4.25}$$

(which defines μ, if you like) so that

$$\cancel{M} x_{\mathrm{cm}} = \frac{1}{2}\mu L^2 = \cancel{M}\frac{L}{2} \tag{4.26}$$

or:

$$\boxed{x_{\mathrm{cm}} = \frac{L}{2}} \tag{4.27}$$

Note that we didn't bother to evaluate a y component because it is surely in the center of a *thin* uniform rod – we've made the problem "one-dimensional" by aligning the rod with the x-axis.

Gee, that was easy. Let's try a harder one, one with *two* dimensions to handle.

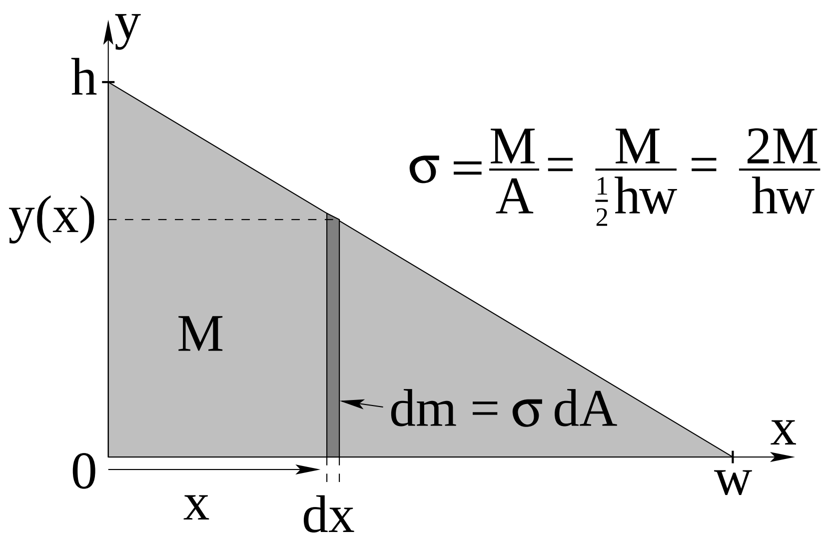


Figure 4.5

Example 4.1.3: Center of mass of a right triangular wedge

Let's do a 2D example. Above is a thin, flat, uniform sheet of metal of mass M in the shape of a triangle with height h and width w. We want to ***find its center of mass***.

Now we need mass per unit *area*, usually assigned the symbol σ (probably for "(s)urface" density) in physics books:

$$\sigma = \frac{M}{A} = \frac{2M}{hw} \tag{4.28}$$

and a suitable differential chunk dA. To avoid "real" multivariate calculus for one more page, we'll do the coordinates x_{cm} and y_{cm} separately and we'll pick *differential strip* of height $y(x)$ and width dx to integrate over, since we can easily figure out the slope-intercept formula for the straight line: $y(x)$ that forms the upper boundary of the triangle:

$$y(x) = h - (h/w)x \tag{4.29}$$

With this, the differential area of the dark grey shaded chunk in the figure above is:

$$dA = y(x)dx = \left(h - \frac{h}{w}x\right)dx \tag{4.30}$$

and its mass is:

$$dm = \sigma\, dA = \frac{2M}{hw}\left(h - \frac{h}{w}x\right)dx \tag{4.31}$$

Now we can *easily* integrate to find x_{cm}:

$$\begin{aligned} x_{cm} &= \frac{1}{M}\int x\ dm = \frac{1}{M}\int_0^w x(\sigma dA) = \frac{1}{\cancel{M}}\frac{2\cancel{M}}{\cancel{h}w}\int_0^w (\cancel{h}x - \frac{\cancel{h}}{w}x^2)dx \\ &= \frac{2}{w}\left(\frac{x^2}{2} - \frac{1}{w}\frac{x^3}{3}\right)\Bigg|_0^w = w - \frac{2}{3}w = \boxed{\frac{w}{3}} \end{aligned} \tag{4.32}$$

We now *could* set up and do a similar integral for y_{cm}, but it is *much simpler* to just ***rename*** $x \Rightarrow y$ and $w \Rightarrow h$) to get:

$$y_{cm} = \boxed{h/3} \tag{4.33}$$

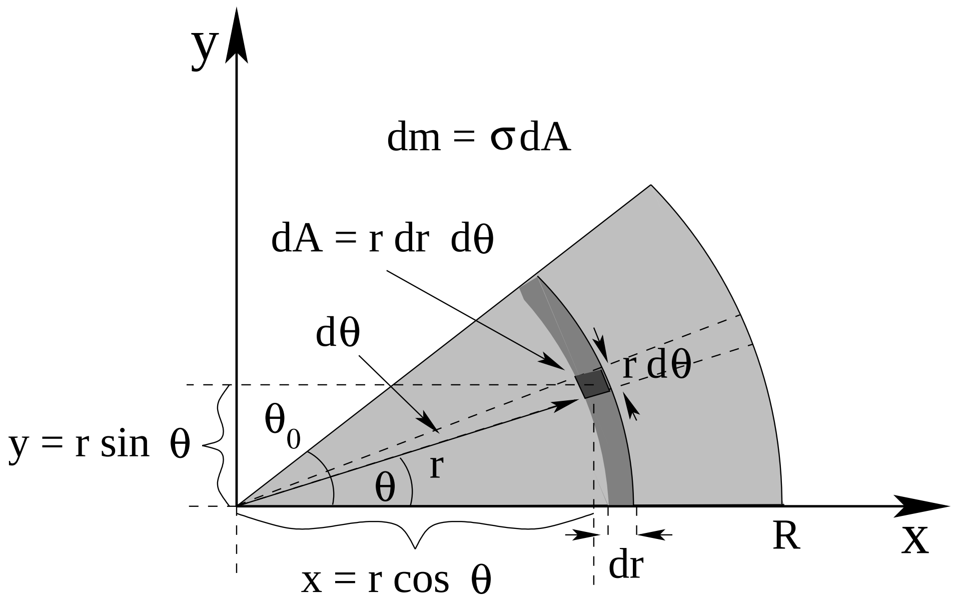


Figure 4.6

Example 4.1.4: Center of mass of a circular wedge

Now for a hard one, a 2D integral in polar coordinates! Let's find the center of mass of a flat circular wedge (a shape like a piece of pie, but very flat) made again of uniformly thick metal or plastic. It has radius R, a given mass M, an angular width of θ_0. We want to find its center of mass in cartesian coordinates, x_{cm} and y_{cm}, and this time we *will* have to do two integrals or else work just as hard with trigonometry to avoid the second one.

Note Well! We integrate over *polar* coordinates, instead of cartesian, because in this coordinate system ***the two integrals over θ and r separate*** and can be done completely independently. This means that we avoid all of the *hard part* of multivariate calculus – really we just use single variable calculus, twice. In cartesian coordinates, integrating inside a circular (or cylindrical, or spherical) domain is a *pain in the ass* and to be avoided if at all possible – it's one reason that you will need to learn cartesian *and* cylindrical (plane polar plus a z axis) *and* spherical polar coordinates in order to truly master this stuff, even if we postpone the harder stuff until later.

It also means that we need to find the *area element* – the standard "differential chunk of area" in polar coordinates. Consider the dark grey rectangle-ish-shaped area in figure 4.6 above (it's a little busy, sorry). It is located at position r, θ in polar coordinates – its x-y coordinates are (obviously) $x = r\cos\theta$, $y = r\sin\theta$. We'll need these shortly. This rectangle is one small segment of width $d\theta$ from the *light* grey strip at radius r with thickness dr. The actual *length along the arc* of this chunk is $r\,d\theta$ the length of the circular arc of angular width $d\theta$ (bounded by the dashed radii from the origin out to R). Therefore, its area is:

$$dA = (r\,d\theta)dr = r\,dr\,d\theta \tag{4.34}$$

The *mass* of this chunk is found from the litany "the mass of the chunk is the mass per unit area times the area of the chunk", or:

$$dm = \sigma dA \tag{4.35}$$

nbut we were not *given* σ, we were given the mass M! Our first chore, therefore, is to find it! Note that:

$$M = \int dm = \int_0^R \int_0^{\theta_0} \sigma dA = \int_0^R \int_0^{\theta_0} \sigma r dr d\theta = \sigma \frac{R^2\theta_0}{2} \tag{4.36}$$

from which (yes!) we can find:

$$\sigma = \frac{2M}{R^2\theta_0} \quad \Rightarrow \quad dm = \frac{2M}{R^2\theta_0} r\,dr\,d\theta \tag{4.37}$$

All that is left is for us to set up the integrals and do them! I'll do x_{cm} for you here, and you should do y_{cm} on your own (do *every step!*) to help you solidify this and remember it. Recall our derived definition:

$$x_{\rm cm} = \frac{1}{M}\int_{\rm wedge} x dm \tag{4.38}$$

Integrating over the entire wedge is *easy* in polar coordinates – the boundaries are $\theta = 0$, $\theta = \theta_0$, $r = 0$, and $r = R$, so we can ***just use these as the limits of two independent integrals!*** If we used cartesian coordinates, we might be able to manage the upper straight line the way we did for a triangle, but the outer circular arc boundary would cause us to pull our hair out and beat our head against a "full course in multivariate calculus" wall. It's not impossible, just extraordinarily unpleasant and a lot more *math* than physics!

Hence we use polar coordinates:

$$\begin{aligned}
x_{\rm cm} &= \frac{1}{M}\int_0^R \int_0^{\theta_0} x dm \\
&= \frac{1}{\cancel{M}}\int_0^R \int_0^{\theta_0} (r\cos\theta)\frac{2\cancel{M}}{R^2\theta_0} r\,dr\,d\theta \\
&= \frac{2}{R^2\theta_0} \times \int_0^R r^2 dr \int_0^{\theta_0} \cos\theta\; d\theta \\
&= \frac{2}{\cancel{R^2}\theta_0} \times \frac{R^{\cancel{3}}}{3} \times \sin\theta_0 = \boxed{\left(\frac{2}{3}\frac{\sin\theta_0}{\theta_0}\right) R}.
\end{aligned} \tag{4.39}$$

Amazingly enough, everything in the parentheses is dimensionless, and R has *correctly* got units of length! Also, if we did the entire disk – $\theta_0 \to 2\pi$, in other words – $\sin 2\pi = 0$, $x_{\rm cm} = 0$, which is *correct!* If we did a half circle, $x_{\rm cm} = 0$ *again* which is correct. We can do one more extremely simple check. For a very *small* angle θ_0:

$$\frac{\sin\theta_0}{\theta_0} \approx 1 \tag{4.40}$$

so that:

$$X_{\rm cm} = \frac{2}{3} R \tag{4.41}$$

This is exactly what we get for a 2D ***right triangle*** (see previous example)! When the angle in the circular wedge is very small, the wedge shape approaches that of a *very* acute right triangle, so as far as we can tell, our answer actually *makes sense!*

Don't forget! I left $y_{\rm cm}$ for you to do on your own. Note that it really only requires you to make *one small change* in *one of the two integrals*, but you should still draw your own picture, work out σ, dA and $dm = \sigma dA$, and then set up and do the double integral on your own. I *do* like to put wedge or arc problems onto quizzes or exams, so be prepared for this!

Example 4.1.5: Breakup of Projectile in Midflight

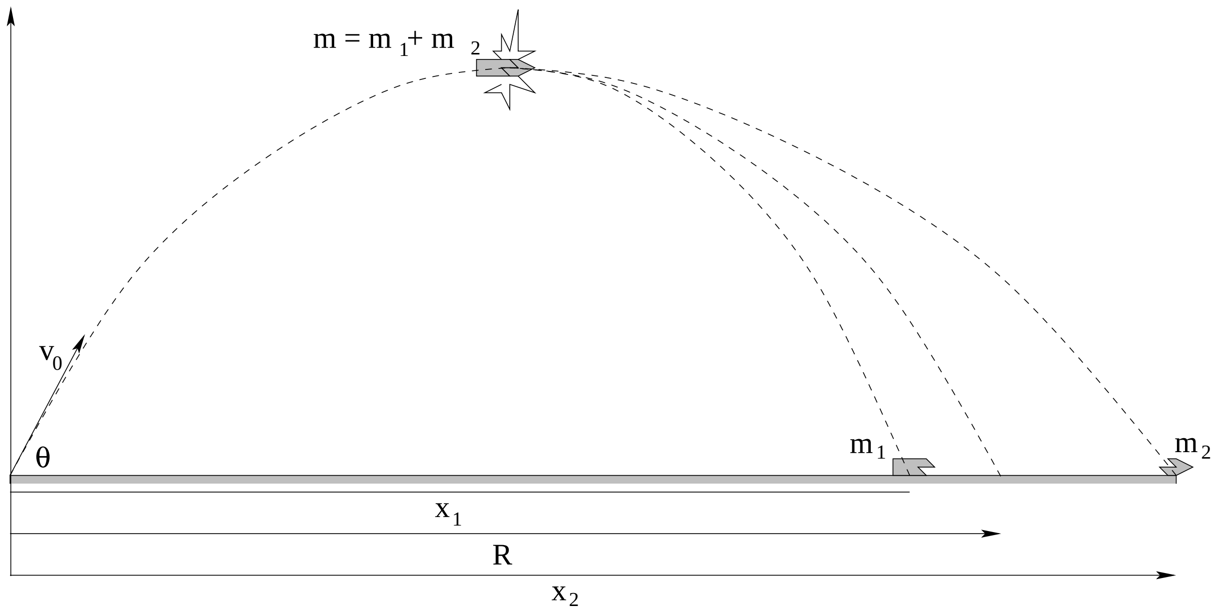


Figure 4.7: A projectile breaks up in midflight. The center of mass follows the original trajectory of the particle, allowing us to predict where one part lands if we know where the other one lands, as long as the explosion exerts no vertical component of force on the two particles.

Suppose that a projectile breaks up horizontally into two pieces of mass m_1 and m_2 in midflight. Given θ, v_0, and x_1, predict x_2.

The idea is: The trajectory of the center of mass obeys Newton's Laws for the entire projectile and lands in the same place that it would have, because no external forces *other* than gravity act. The projectile breaks up horizontally, which means that both pieces will land at the same time, with the center of mass in between them. We thus need to find the point where the center of mass would have landed, and solve the equation for the center of mass in terms of the two places the projectile fragments land for one, given the other. Thus:

Find R. As usual:

$$y = (v_0 \sin\theta)t - \frac{1}{2}gt^2 \tag{4.42}$$

$$t_R(v_0 \sin\theta - \frac{1}{2}gt_R) = 0 \tag{4.43}$$

$$t_R = \frac{2v_0 \sin\theta)}{g} \tag{4.44}$$

$$R = (v_0 \cos\theta)t_R = \frac{2v_0^2 \sin\theta\cos\theta}{g}. \tag{4.45}$$

R is the position of the center of mass. We write the equation making it so:

$$m_1x_1 + m_2x_2 = (m_1 + m_2)R \tag{4.46}$$

and solve for the unknown x_2.

$$x_2 = \frac{(m_1 + m_2)R - m_1x_1}{m_2} \tag{4.47}$$

From this example, we see that it is sometimes easiest to solve a problem by separating the motion *of* the center of mass of a system from the motion *in* a reference frame that "rides along" with the center of mass. The price we may have to pay for this convenience is the appearance of pseudoforces in this frame if it happens to be accelerating, but in many cases it will *not* be accelerating, or the acceleration will be so small that the pseudoforces can be neglected compared to the much larger forces of interest acting within the frame. We call this (at least approximately) inertial reference frame the ***Center of Mass Frame*** and will discuss and define it in a few more pages.

First, however, we need to define an extremely useful concept in physics, that of ***momentum***, and discuss the closely related concept of ***impulse*** and the ***impulse approximation*** that permits us to treat the center of mass frame as being approximately inertial in many problems even when it is accelerating.

4.1.3: Center of Mass of Two or More Extended Objects

So far, all we've done is find the center of mass of a collection of *particles*, or used calculus in the limit that dm behaves like a "particle" to do the sums. What happens when we have more than one *continuous* object, say, two flat semicircles cut out of steel and separated by the distance R, or (for later) a grandfather clock pendulum made up of a rod of length L with a heavy disk at the end, or a disk with a circular *hole* cut out? Those all seem as though they'd be almost impossible to do via direct integration, as we can't easily pick a *single* coordinate system that works well for *both* objects, or the one object with the hole. And things will only get worse with more objects, more holes!

Fortunately, there is an easily derived ***theorem*** that will let us solve problems like this extremely easily, and explain something we might have guessed as well. Let's do the heavy lifting with the discrete sum, and then we'll just convert the result into integral(s) form.

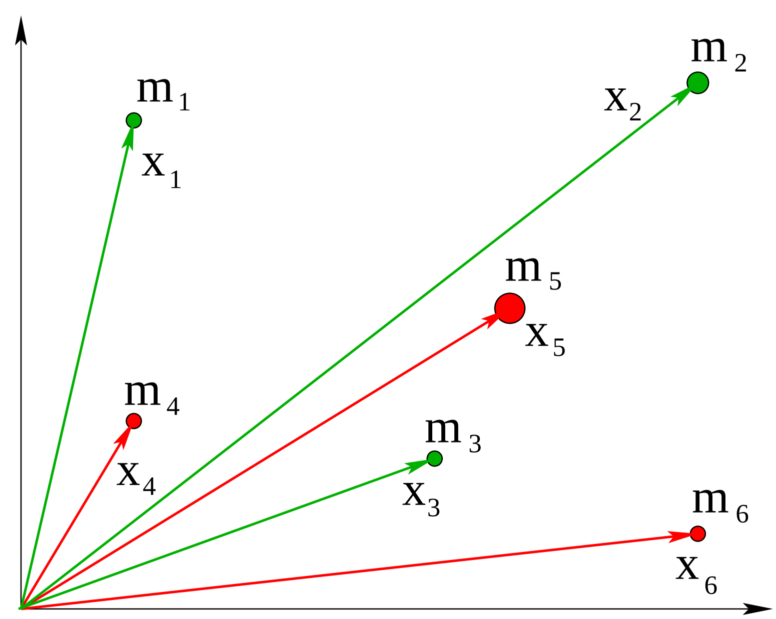


Figure 4.8: $O = 6$ particles, 3 green and 3 red, scattered arbitrarily in space to help you visualize the derivation below.

Suppose we have O particles scattered around in space, and want to find the center of mass of this system. However, M of the particles are (say) green, and N of them are (say) red, so that $O = N + M$. Then:

$$\vec{\boldsymbol{x}}_{\rm cm} = \frac{1}{M_{\rm tot}} \sum_{i=1}^{O} m_i \vec{\boldsymbol{x}}_i \tag{4.48}$$

as usual.

We can also *independently* find the center of mass of the green objects and red objects, where we'll sort out the indices so all of the green object indices precede the red object indices to keep the partitioning simple:

$$\vec{\boldsymbol{x}}_{{\rm cm},g} = \frac{1}{M_{{\rm tot},g}} \sum_{i=1}^{M} m_i \vec{\boldsymbol{x}}_i \qquad \text{and} \qquad \vec{\boldsymbol{x}}_{{\rm cm},r} = \frac{1}{M_{{\rm tot},r}} \sum_{i=M+1}^{O} m_i \vec{\boldsymbol{x}}_i \tag{4.49}$$

Hopefully it is clear that:

$$M_{{\rm tot},g} = \sum_{i=1}^{M} m_i \qquad M_{{\rm tot},r} = \sum_{i=M+1}^{O} m_i \qquad M_{\rm tot} = M_{{\rm tot},g} + M_{{\rm tot},r}$$

Let's rearrange these last two equations into:

$$M_{{\rm tot},g} \vec{\boldsymbol{x}}_{{\rm cm},g} = \sum_{i=1}^{M} m_i \vec{\boldsymbol{x}}_i \qquad \text{and} \qquad M_{{\rm tot},r} \vec{\boldsymbol{x}}_{{\rm cm},r} = \sum_{i=M+1}^{O} m_i \vec{\boldsymbol{x}}_i \tag{4.50}$$

then ***add them***:

$$\begin{aligned} M_{{\rm tot},g} \vec{\boldsymbol{x}}_{{\rm cm},g} + M_{{\rm tot},r} \vec{\boldsymbol{x}}_{{\rm cm},g} &= \sum_{i=1}^{M} m_i \vec{\boldsymbol{x}}_i + \sum_{i=M+1}^{O} m_i \vec{\boldsymbol{x}}_i \\ &= \sum_{i=1}^{O} m_i \vec{\boldsymbol{x}}_i \\ &= M_{\rm tot} \vec{\boldsymbol{x}}_{\rm cm} \end{aligned}$$

or

$$M_{\rm tot} \vec{\boldsymbol{x}}_{\rm cm} = M_{{\rm tot},g} \vec{\boldsymbol{x}}_{{\rm cm},g} + M_{{\rm tot},r} \vec{\boldsymbol{x}}_{{\rm cm},g} \tag{4.51}$$

Note that the sum is just $M_{\rm tot} \vec{\boldsymbol{x}}_{\rm cm}$! If we divide both sides by $M_{\rm tot}$, then, we find that:

$$\vec{\boldsymbol{x}}_{\rm cm} = \frac{1}{M_{\rm tot}} = M_{{\rm tot},g} \vec{\boldsymbol{x}}_{{\rm cm},g} + M_{{\rm tot},r} \vec{\boldsymbol{x}}_{{\rm cm},g} \tag{4.52}$$

Clearly it doesn't matter how many ways we partition the particles or how many colors we use. What we have shown is that:

> A collection of particles behaves just like a particle with the total mass of the collection located at the center of mass of the collection for the purpose of finding the *total* center of mass of many such collections!

This is now ***proven***, as a *theorem*.

It also clearly doesn't matter if the collections are discrete particles or extended continuous objects. Imagine that the green masses are all stuck together into a green *lime* (extended object) and all of the red masses are stuck together into a red *apple!* Obviously the center of mass of the lime and apple themselves can be found by integration in any coordinate frame we find convenient, and we can then find the center of mass of the combined lime-apple system *by pretending that they are point masses located at their respective centers of mass!*

Our theorem isn't limited to only *two* extended objects, or collections of point like objects – we could have had three, or twenty, or leventy-zillion objects! Again, this reinforces the idea that we can indeed treat pens, cars, baseballs, and clouds of interstellar gas like *objects* as long as we locate those objects at their respective centers of mass[113]

What we have shown is that the center of mass, as an idea, is in some sense *recursive*. Once we find the center of mass of an object, we can use that position as the position of a "particle" with the same total mass to calculate the center of mass of a *new* system made up of the object and *other* objects/particles, combined.

This is both extraordinary and expected. We have from the beginning treated things like baseballs like particles, so it really shouldn't be any surprise that the center of mass of two identical baseballs is halfway in between their individual centers of mass, exactly as though the baseballs were true particles with the mass of a baseball located precisely at those centers. However, it works just as well to find the mutual center of mass of a baseball and a *bat*[114], where the bat can have any mass or orientation relative to the baseball as long as we know its mass and where its center of mass is. It means that we can reduce the problem of finding the center of mass of the solar system to determining the masses and positions of the centers of all of the roughly spherical planets and doing a simple sum, instead of trying to formulate a single integral that somehow covers all of the planets and asteroids and so on. It means that you can now do new kinds of *problems* with the simple discrete sum rule *conceptually*, without using calculus at all, as long as you can make a pretty good guess as to the center of mass of nice, symmetric objects! Let's try a couple.

[113]And, of course, carefully account for ever so much physics we haven't learned yet, like torque and rotation, viscosity, long range forces like gravity and electromagnetic radiation etc. But the idea itself is still sound and the theorem still holds, even with all of these.

[114]Yes, I meant the kind of bat you hit baseballs with, but what our theorems shows is that it works just as well for a baseball and a bat of the *furry, night-flying* kind.

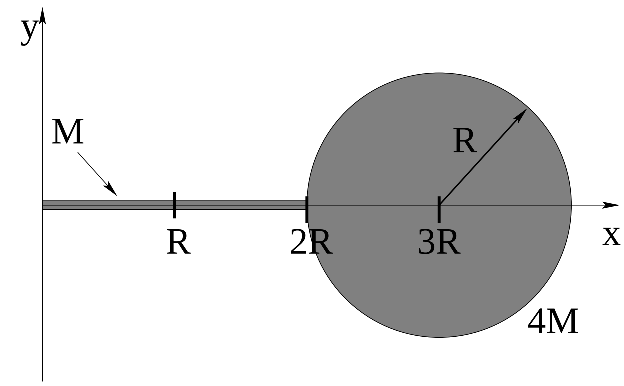


Figure 4.9

Example 4.1.6: Rod and Ball

In figure 4.9 a uniform rod of mass M and length $2R$ is connected to a uniform spherical ball of mass $4M$ and radius R as shown. We'd like to find the center of mass of this system.

Now it is easy! We don't *have* to actually do the integral over a rod – we did it above and besides, we *guessed* that it would be *in the middle of the uniform rod*, that is, at $(x = R, y = 0)$ on the coordinate frame drawn above. Similarly, it is *obvious* that the center of mass of the sphere is at $(x = 3R, y = 0)$ – something that it would be a bit of a chore to prove even using separable spherical polar integrals and calculus! Hence:

$$x_{\rm cm} = \frac{1}{M + 4M}\left(M \times R + 4M \times 3R\right) = \boxed{\frac{13}{5}R} \qquad \text{(and, of course)} \qquad y_{\rm cm} = 0 \tag{4.53}$$

This is actually inside of the sphere!

How about another?

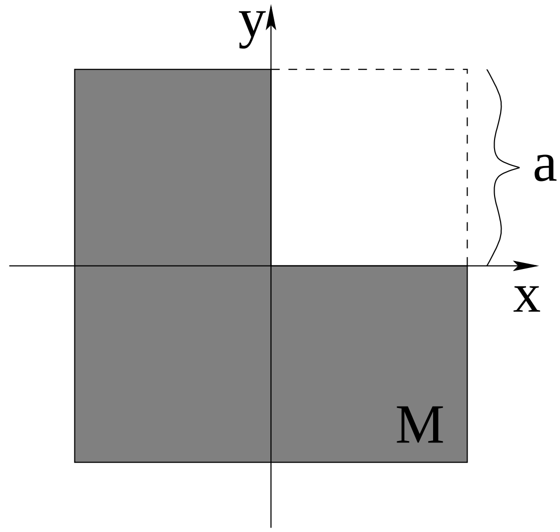


Figure 4.10

Example 4.1.7: Square Missing a Square Corner

In figure 4.10 above, your choice of a flat (2D) square with sides of length $2a$ missing a square with side a out of one corner, or three squares with side a arranged as drawn. Either way, the actual L-sha[ed object has mass M and we want to find the center of mass of the object.

At this point, you should be able to see that *one* way to do it is to locate the centers of mass of the three small sub-squares, give each of them a mass of $M/3$, and treat this as a three particle problem. We'll do it this way below, but first I'm going to demonstrate a slightly different way to do it. It will give us the same answer, but it is a niftier approach and will allow you to do problems that *can't* easily be done just by summing over chunks of the actual object.

I'm going to imagine that the system in question is a *complete* square with sides of length $2a$ and total mass $4M/3$ – which now *obviously* has its center of mass at the origin of the coordinate system drawn. The second object is then a small square of imaginary *anti-mass* (not antimatter per se, as that would make a spectacular explosion and we can't have that). The small square has sides of length a and mass $-M/3$ and a center of mass located at *its* geometric center at $(a/2, a/2)$. Clearly if we lay the small square down in the upper corner, its mass will *cancel* the mass of the big square in that corner to produce a final distribution just like the picture! This trick is worth remembering for almost any "nice" shape *with holes*, places where the mass is missing.

Using this trick we can find the center of mass as follows:

$$x_{\rm cm} = \frac{1}{M}\left(\frac{4}{3}M \times 0 + \left(-\frac{M}{3}\right) \times \frac{1}{2}a\right) = \boxed{-\frac{1}{6}a} \tag{4.54}$$

$$y_{\rm cm} = \boxed{-\frac{1}{6}a} \tag{4.55}$$

I avoided a lot of work by appealing to good-old *symmetry* and observing that since the center of mass has to be on the diagonal line of mirror symmetry of the system, $y_{\rm cm} = x_{\rm cm}$.

For fun, let's do this the *other* way (still using this theorem) – finding the center of mass of the three physical squares that make up the total mass M:

$$x_{\rm cm} = \frac{1}{M}\left(\frac{M}{3} \times -\frac{a}{2} + \frac{M}{3} \times -\frac{a}{2} + \frac{M}{3} \times \frac{a}{2}\right) = -\frac{1}{6}a \tag{4.56}$$

$$y_{\rm cm} = \boxed{-\frac{1}{y}a} \tag{4.57}$$

Same answer (still using symmetry to avoid explicitly solving for $y_{\rm cm} = x_{\rm cm}$), and arguably a *little* more algebra along the way.

4.2: Momentum

Momentum is a useful idea that follows naturally from our decision to treat collections as objects. It is a way of combining the mass (which is a characteristic of the object) with the velocity of the object. We **define** the **momentum** to be:

$$\vec{\boldsymbol{p}} = m\vec{\boldsymbol{v}} \tag{4.58}$$

Thus (since the mass of an object is generally constant):

$$\vec{\boldsymbol{F}} = m\vec{\boldsymbol{a}} = m\frac{d\vec{\boldsymbol{v}}}{dt} = \frac{d}{dt}(m\vec{\boldsymbol{v}}) = \frac{d\vec{\boldsymbol{p}}}{dt} \tag{4.59}$$

is another way of writing Newton's second law. In fact, this is the way Newton actually **wrote** Newton's second law – he did not say "$\vec{\boldsymbol{F}} = m\vec{\boldsymbol{a}}$" the way we have been reciting. We emphasize this connection because it makes the path to solving for the trajectories of constant mass particles a bit easier, not because things really make more sense that way.

Note that there exist systems (like rocket ships, cars, etc.) where the mass is **not** constant. As the rocket rises, its thrust (the force exerted by its exhaust) can be constant, but it continually gets lighter as it burns fuel. Newton's second law (expressed as $\vec{\boldsymbol{F}} = m\vec{\boldsymbol{a}}$) **does** tell us what to do in this case – but only if we treat each little bit of burned and exhausted gas as a "particle", which is a pain. On the other hand, Newton's second law expressed as $\vec{\boldsymbol{F}} = \frac{d\vec{\boldsymbol{p}}}{dt}$ still works fine and makes perfect sense – it simultaneously describes the loss of mass and the increase of velocity as a function of the mass correctly.

Clearly we can repeat our previous argument for the sum of the momenta of a collection of particles:

$$\vec{\boldsymbol{P}}_{\rm tot} = \sum_i \vec{\boldsymbol{p}}_i = \sum_i m\vec{\boldsymbol{v}}_i \tag{4.60}$$

so that

$$\frac{d\vec{\boldsymbol{P}}_{\rm tot}}{dt} = \sum_i \frac{d\vec{\boldsymbol{p}}_i}{dt} = \sum_i \vec{\boldsymbol{F}}_i = \vec{\boldsymbol{F}}_{\rm tot} \tag{4.61}$$

Differentiating our expression for the position of the center of mass above, we also get:

$$\frac{d\sum_i m_i\vec{\boldsymbol{x}}_i}{dt} = \sum_i m_i\frac{d\vec{\boldsymbol{x}}_i}{dt} = \sum_i \vec{\boldsymbol{p}}_i = \vec{\boldsymbol{P}}_{\rm tot} = M_{\rm tot}\vec{\boldsymbol{v}}_{\rm cm} \tag{4.62}$$

4.2.1: The Law of Conservation of Momentum

We are now in a position to state and trivially prove the **Law of Conservation of Momentum**. It reads[115]:

> ***If and only if*** the total external force acting on a system is zero, ***then*** the total momentum of a system (of particles) is a constant vector.

[115]The "if and only if" bit, recall, means that if the total momentum of a system is a constant vector, it *also* implies that the total force acting on it is zero, there is no other way that this condition can come about.

You are welcome to learn this in its more succinct algebraic form:

$$\text{If and only if } \vec{F}_{\rm tot} = 0 \text{ then } \vec{P}_{\rm tot} = \vec{P}_{\rm initial} = \vec{P}_{\rm final} = \text{ a constant vector.} \tag{4.63}$$

Please learn this law *exactly* as it is written here. The condition $\vec{F}_{\rm tot} = 0$ is *essential* – otherwise, as you can see, $\vec{F}_{\rm tot} = \frac{d\vec{P}_{\rm tot}}{dt}$!

The proof is almost a one-liner at this point:

$$\vec{F}_{\rm tot} = \sum_i \vec{F}_i = 0 \tag{4.64}$$

implies

$$\frac{d\vec{P}_{\rm tot}}{dt} = 0 \tag{4.65}$$

so that $\vec{P}_{\rm tot}$ is a constant if the forces all sum to zero. This is not quite enough. We need to note that for the **internal** forces (between the ith and jth particles in the system, for example) from Newton's third law we get:

$$\vec{F}_{ij} = -\vec{F}_{ji} \tag{4.66}$$

so that

$$\vec{F}_{ij} + \vec{F}_{ji} = 0 \tag{4.67}$$

pairwise, between **every** pair of particles in the system. That is, although internal forces may not be zero (and generally are not, in fact) the changes the cause in the momentum of the system cancel. We can thus subtract:

$$\vec{F}_{\rm internal} = \sum_{i,j} \vec{F}_{ij} = 0 \tag{4.68}$$

from $\vec{F}_{\rm tot} = \vec{F}_{\rm external} + \vec{F}_{\rm internal}$ to get:

$$\vec{F}_{\rm external} = \frac{d\vec{P}_{\rm tot}}{dt} = 0 \tag{4.69}$$

and the total momentum must be a constant (vector).

This can be thought of as the "bootstrap law" – *You cannot lift yourself up by your own bootstraps!* No matter what force one part of you exerts on another, those internal forces can never alter the velocity of your center of mass or (equivalently) your total momentum, nor can they overcome or even alter any net external force (such as gravity) to lift you up.

As we shall see, the idea of momentum and its conservation greatly simplify doing a wide range of problems, just like energy and its conservation did in the last chapter. It is especially useful in understanding what happens when one object *collides* with another object.

Evaluating the dynamics and kinetics of microscopic collisions (between, e.g. electrons, protons, neutrons and targets such as atoms or nuclei) is a big part of contemporary physics – so big that we call it by a special name: **Scattering Theory**[116] . The idea is to take some initial

[116]Wikipedia: http://www.wikipedia.org/wiki/Scattering Theory. This link is mostly for more advanced students, e.g. physics majors, but future radiologists might want to look it over as well as it is the basis for a whole lot of radiology...

(presumed known) state of an about-to-collide "system", to let it collide, and to either infer from the observed scattering something about the nature of the force that acted during the collision, or to predict, from the measured final state of some of the particles, the final state of the rest.

Sound confusing? It's not, really, but it can be **complicated** because there are lots of things that might make up an initial and final state. In this class we have humbler goals – we will be content simply understanding what happens when *macroscopic* objects like cars or billiard[117] balls collide, where (as we will see) momentum conservation plays an enormous role. This is still the first step (for physics majors or future radiologists) in understanding more advanced scattering theory but it provides a lot of direct insight into everyday experience and things like car safety and why a straight on shot in pool often stops one ball cold while the other continues on with the original ball's velocity.

In order to be able to use momentum conservation in a collision, however, no *external* force can act on the colliding objects during the collision. This is almost never going to *precisely* be the case, so we will have to idealize by assuming that a "collision" (as opposed to a more general and leisurely force interaction) involves forces that are zero right up to where the collision starts, spike up to very large values (generally much larger than the sum of the other forces acting on the system at the time) and then drop quickly back to zero, being *non*-zero only in a very short time interval Δt.

In this idealization, collisions will (by assumption) take place so fast that any other external forces cannot significantly alter the momentum of the participants during the time Δt. This is called the ***impulse approximation***. With the impulse approximation, we can neglect all other external forces (if any are present) and use ***momentum conservation*** as a key principle while analyzing or solving collisions. *All* collision problems solved in this course should be solved using the impulse approximation. Let's see just what "impulse" is, and how it can be used to help solve collision problems and understand things like the forces exerted on an object by a fluid that is in contact with it.

4.3: Impulse

Let us imagine a typical collision: one pool ball approaches and strikes another, causing both balls to recoil from the collision in some (probably different) directions and at different speeds. Before they collide, they are widely separated and exert no force on one another. As the surfaces of the two (hard) balls come into contact, they "suddenly" exert relatively large, relatively violent, equal and opposite forces on each other over a relatively short time, and then the force between the objects once again drops to zero as they either bounce apart or stick together and move with a common velocity. "Relatively" here in all cases means ***compared to all other forces acting on the system during the collision*** in the event that those forces are

[117]Wikipedia: http://www.wikipedia.org/wiki/Billiards. It is always dangerous to assume the every student has had any given experience or knows the same games or was raised in the same culture as the author/teacher, especially nowadays when a significant fraction of *my* students, at least, come from other countries and cultures, and when this book is in use by students all over the world outside of my own classroom, so I provide this (and various other) links. In this case, as you will see, billiards or "pool" is a game played on a table where the players try to knock balls in holes by poking one ball (the "cue ball") with a stick to drive another identically sized ball into a hole. Since the balls are very hard and perfectly spherical, the game is an excellent model for two-dimensional elastic collisions.

not actually zero.

For example, when skidding cars collide, the collision occurs so fast that even though kinetic friction is acting, it makes an *ignorable* change in the momentum of the cars during the collision *compared to* the total change of momentum of each car due to the collision force. When pool balls collide we can similarly ignore the drag force of the air or frictional force exerted by the table's felt lining for the tiny time they are in actual contact. When a bullet embeds itself in a block, it does so so rapidly that we can ignore the friction of the table on which the block sits. Idealizing and ignoring e.g. friction, gravity, drag forces in situations such as this is known as ***the impulse approximation***, and it greatly simplifies the treatment of collisions.

Note that we will frequently not know the detailed functional form of the collision force, $\vec{F}_{\rm coll}(t)$ nor the *precise* amount of time Δt in any of these cases. The "crumpling" of cars as they collide is a very complicated process and exerts a *completely unique force* any time such a collision occurs – no two car collisions are exactly alike. Pool balls probably do exert a much more reproducible and understandable force on one another, one that we we *could* model if we were advanced physicists or engineers working for a company that made billiard tables and balls and our livelihoods depended on it but we're not and it doesn't. Bullets embedding themselves in blocks again do so with a force that is different every time that we can never *precisely* measure, predict, or replicate.

In all cases, although the *details* of the interaction force are unknown (or even unknowable in any meaningful way), we can obtain or estimate or measure some *approximate* things about the forces in any given collision situation. In particular we can put reasonable limits on Δt and make 'before' and 'after' measurements that permit us to compute the *average* force exerted over this time.

Let us begin, then, by defining the average force over the (short) time Δt of any given collision, assuming that we *did* know $\vec{F} = \vec{F}_{21}(t)$, the force one object (say m_1) exerts on the other object (m_2). The magnitude of such a force (one perhaps appropriate to the collision of pool balls) is sketched below in figure 4.11 where for simplicity we assume that the force acts only along the line of contact and is hence effectively one dimensional in this direction[118].

The time average of this force is computed the same way the time average of any other time-dependent quantity might be:

$$\vec{F}_{\rm avg} = \frac{1}{\Delta t} \int_0^{\Delta t} \vec{F}(t)\ dt \tag{4.70}$$

We can evaluate the integral using Newton's Second Law expressed in terms of momentum:

$$\vec{F}(t) = \frac{d\vec{p}}{dt} \tag{4.71}$$

so that (multiplying out by dt and integrating):

$$\vec{p}_{2f} - \vec{p}_{2i} = \Delta\vec{p}_2 = \int_0^{\Delta t} \vec{F}(t)\ dt \tag{4.72}$$

[118]This is, as anyone who plays pool knows from experience, an excellent assumption and is in fact how one most generally "aims" the targeted ball (neglecting all of the various fancy tricks that can alter this assumption and the outcome).

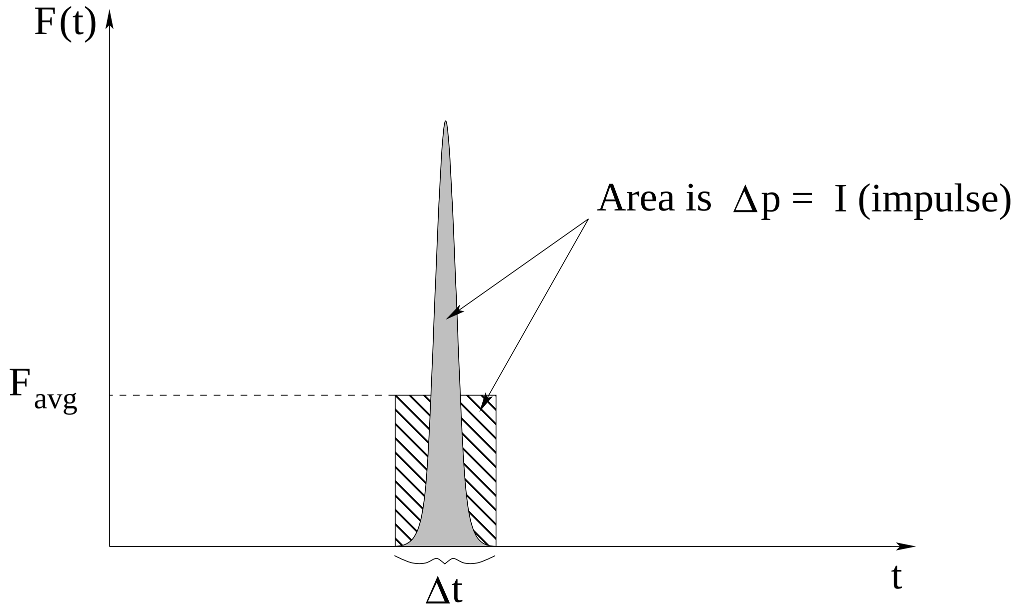


Figure 4.11: A "typical" collision force that might be exerted by the cue ball on the eight ball in a game of pool, approximately along the line connecting the two ball centers. In this case we would expect a fairly symmetric force as the two balls briefly deform at the point of contact. The *time* of contact Δt has been measured to be on the order of a tenth of a millisecond for colliding pool balls.

This is the total vector momentum *change* of the second object during the collision and is also the *area underneath the* $\vec{F}(t)$ *curve* (for each component of a general force – in the figure above we assume that the force only points along one direction over the entire collision and the change in the momentum component in this direction is then the area under the drawn curve). Note that the momentum change of the *first* ball is equal and opposite. From Newton's Third Law, $\vec{F}_{12}(t) = -\vec{F}_{21}(t) = \vec{F}$ and:

$$\vec{p}_{1f} - \vec{p}_{1i} = \Delta\vec{p}_1 = -\int_0^{\Delta t} \vec{F}(t)\, dt = -\Delta\vec{p}_2 \tag{4.73}$$

The integral of a force $\vec{F}$ over an interval of time is called the ***impulse***[119] imparted by the force

$$\vec{I} = \int_{t_1}^{t_2} \vec{F}(t)\, dt = \int_{t_1}^{t_2} \frac{d\vec{p}}{dt}\, dt = \int_{p_1}^{p_2} d\vec{p} = \vec{p}_2 - \vec{p}_1 = \Delta\vec{p} \tag{4.74}$$

This proves that the (vector) impulse is equal to the (vector) change in momentum over the same time interval, a result known as the ***impulse-momentum theorem***. From our point of view, the impulse is just the momentum transferred *between* two objects in a collision in such a way that the *total* momentum of the two is unchanged.

Returning to the average force, we see that the average force in terms of the impulse is just:

$$\vec{F}_{\rm avg} = \frac{\vec{I}}{\Delta t} = \frac{\Delta p}{\Delta t} = \frac{\vec{p}_f - \vec{p}_i}{\Delta t} \tag{4.75}$$

[119]Wikipedia: http://www.wikipedia.org/wiki/Impulse (physics).

If you refer again to figure 4.11 you can see that the area under $F_{\rm avg}$ is equal the area under the actual force curve. This makes the average force relatively simple to compute or estimate any time you know the *change in momentum* produced by a collision and have a way of measuring or assigning an effective or average time Δt per collision.

Example 4.3.1: Average Force Driving a Golf Ball

A golf ball leaves a 1 wood at a speed of (say) 70 meters/second (this is a reasonable number – the world record as of this writing is 90 meters/second). It has a mass of 45 grams. The time of contact has been measured to be $\Delta t = 0.0005$ seconds (very similar to a collision between pool balls). What is the magnitude of the average force that acts on the golf ball during this "collision"?

This one is easy:

$$F_{\rm avg} = \frac{I}{\Delta t} = \frac{mv_f - m(0)}{\Delta t} = \frac{3.15}{0.0005} = 6300 \quad \text{Newtons} \tag{4.76}$$

Since I personally have a mass conveniently (if embarrassingly) near 100 kg and therefore weigh 1000 Newtons, the golf club exerts an *average* force of 6.3 times my weight, call it 3/4 of a ton. The *peak* force, assuming an impact shape for $F(t)$ not unlike that pictured above is as much as two English tons (call it 17400 Newtons).

Note Well! Impulse is related to a whole spectrum of conceptual mistakes students often make! Here's an example that many students would get wrong *before* they take mechanics and that ***no student should ever get wrong after they take mechanics!*** But many do. Try not to be one of them...

Example 4.3.2: Force, Impulse and Momentum for Windshield and Bug

There's a song by Mary Chapin Carpenter called "The Bug" with the refrain:

> Sometimes you're the windshield,
> Sometimes you're the bug...

In a collision between (say) the windshield of a large, heavily laden pickup truck and a teensy little yellowjacket wasp, answer the following qualitative/conceptual questions:

a) Which exerts a larger (magnitude) *force* on the other during the collision?

b) Which changes the magnitude of its *momentum* more during the collision?

c) Which changes the magnitude of its *velocity* more during the collision?

Think about it for a moment, answer all three in your mind. Now, compare it to the *correct answers* below[120]. If you did not get *all three perfectly correct* then go over this whole chapter until you do – you may want to discuss this with your favorite instructor as well.

4.3.1: The Impulse Approximation

When we analyze actual collisions in the real world, it will almost never be the case that there are no external forces acting on the two colliding objects during the collision process. If we hit a baseball with a bat, if two cars collide, if we slide two air-cushioned disks along a tabletop so that they bounce off of each other, gravity, friction, drag forces are often present. Yet we will, in this textbook, uniformly assume that these forces are *irrelevant* during the collision.

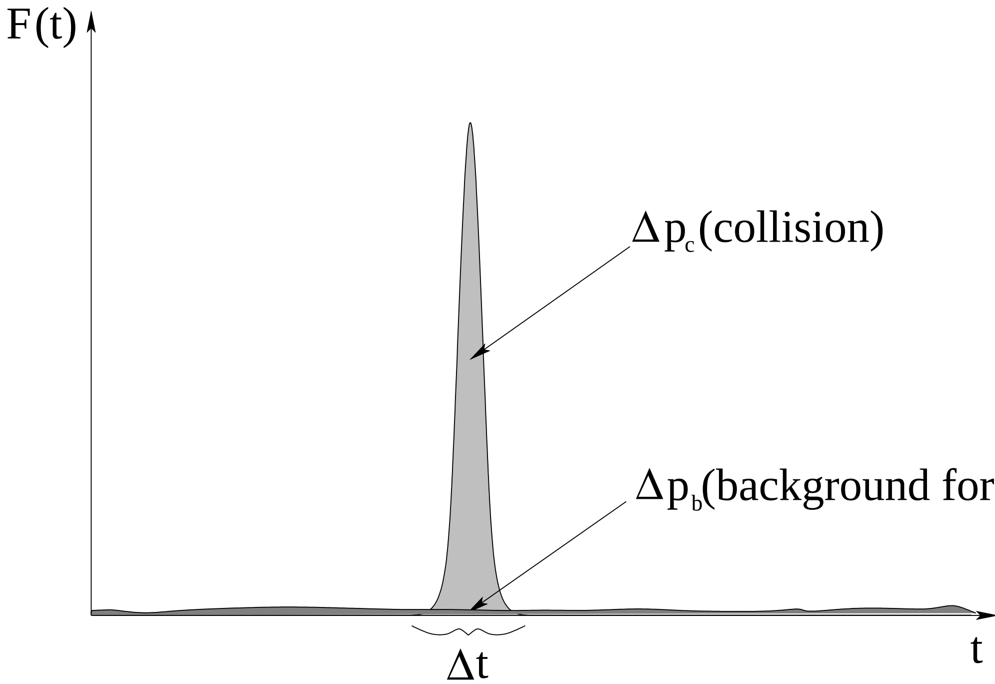


Figure 4.12: Impulse forces for a collision where typical *external* forces such as gravity or friction or drag forces are also present.

Let's see why (and when!) we can get away with this. Figure 4.12 shows a typical collision force (as before) for a collision, but this time shows some external force acting on the mass *at the same time*. This force might be varying friction and drag forces as a car brakes to try to avoid a collision on a bumpy road, for example. Those forces may be large, but in general they are *very small* compared to the peak, or average, collision force between two cars. To put it in perpective, in the example above we estimated that the average force between a golf ball and a golf club is over 6000 newtons during the collision – around *six times my* (substantial)

[120]Put here so you can't see them while you are thinking so easily. The force exerted by the truck on the wasp is ***exactly the same*** as the force exerted by the wasp on the truck (Newton's Third Law!). The magnitude of the momentum (or impulse) transferred from the wasp to the truck is ***exactly the same*** as the magnitude of the momentum transferred from the truck to the wasp. However, the velocity of the truck ***does not measurably change*** (for the probable impulse transferred from any normal non-Mothra-scale wasp) while the wasp (as we will see below) bounces off going roughly twice the speed of the truck...

weight. In contrast, the golf ball itself weighs much less than a newton, and the drag force and friction force between the golf ball and the tee are a tiny fraction of that.

If anything, the background forces in this figure are *highly exaggerated* for a typical collision, compared to the scale of the actual collision force!

The change in momentum resulting from the background force is the area underneath its curve, just as the change in momentum resulting from the collision force alone is the area under the collision force curve.

Over macroscopic time – over seconds, for example – gravity and drag forces and friction can make a significant contribution to the change in momentum of an object. A braking car slows down. A golf ball soars through the air in a gravitational trajectory modified by drag forces. But ***during the collision time Δt they are negligible***, in the specific sense that:

$$\Delta\vec{p} = \Delta\vec{p}_c + \Delta\vec{p}_b \approx \Delta\vec{p}_c \tag{4.77}$$

(for just one mass) over that time only. Since the collision force is an *internal* force between the two colliding objects, it cancels for the system making the momentum change of the system during the collision approximately zero.

We call this approximation $\Delta\vec{p} \approx \Delta\vec{p}_c$ (neglecting the change of momentum resulting from background external forces during the collision) the ***impulse approximation*** and we will *always* assume that it is valid in the problems we solve in this course. It justifies treating the center of mass reference frame (discussed in the next section) as an inertial reference frame *even when technically it is not* for the purpose of analyzing a collision or explosion.

It is, however, useful to have an understanding of when this approximation might *fail*. In a nutshell, it will fail for collisions that take place over a long enough time Δt that the external forces produce a change of momentum that is *not* negligibly small compared to the momentum exchange between the colliding particles, so that the total momentum before the collision is *not* approximately equal to the total momentum after the collision.

This can happen because the external forces are unusually large (comparable to the collision force), or because the collision force is unusually small (comparable to the external force), or because the collision force acts over a *long time* Δt so that the external forces have time to build up a significant $\Delta\vec{p}$ for the system. None of these circumstances are typical, however, although we can imagine setting up an problem where it is true – a collision between two masses sliding on a rough table during the collision where the collision force is caused by a weak spring (a variant of a homework problem, in other words). We will consider this sort of problem (which is considerably more difficult to solve) to be beyond the scope of this course, although it is not beyond the scope of what the concepts of this course would permit you to set up and solve if your life or job depended on it.

4.3.2: Impulse, Fluids, and Pressure

Another valuable use of impulse is when we have *many* objects colliding with something – so many that even though *each* collision takes only a short time Δt, there are so many collisions that they exert a nearly continuous force on the object. This is critical to understanding the notion of *pressure exerted by a fluid*, because microscopically the fluid is just a lot of very small particles that are constantly colliding with a surface and thereby transferring momentum to it, so many that they exert a nearly continuous and smooth force on it that is the *average* force exerted per particle times the number of particles that collide. In this case Δt is conveniently considered to be the inverse of the *rate* (number per second) with which the fluid particles collide with a section of the surface.

To give you a very crude idea of how this works, let's review a small piece of the kinetic theory of gases. Suppose you have a cube with sides of length L containing N molecules of a gas. We'll imagine that all of the molecules have a mass m and an average speed in the x direction of v_x, with (on average) one half going left and one half going right at any given time.

In order to be in *equilibrium* (so v_x doesn't change) the change in momentum of any molecule that hits, say, the right hand wall perpendicular to x is $\Delta p_x = 2mv_x$. This is the *impulse* transmitted to the wall per molecular collision. To find the total impulse in the time Δt, one must multiply this by one half the number of molecules in in a volume $L^2 v_x \Delta t$. That is,

$$\Delta p_{\rm tot} = \frac{1}{2}\left(\frac{N}{L^3}\right) L^2 v_x \Delta t\,(2mv_x) \tag{4.78}$$

Let's call the volume of the box $L^3 = V$ and the area of the wall receiving the impulse $L^2 = A$. We combine the pieces to get:

$$P = \frac{F_{\rm avg}}{A} = \frac{\Delta p_{\rm tot}}{A\ \Delta t} = \left(\frac{N}{V}\right)\left(\frac{1}{2}mv_x^2\right) = \left(\frac{N}{V}\right) K_{x,{\rm avg}} \tag{4.79}$$

where the average force per unit area applied to the wall is the ***pressure***, which has SI units of Newtons/meter^2 or ***Pascals***.

If we add a result called the ***equipartition theorem***[121] :

$$K_{x,{\rm avg}} = \frac{1}{2}mv_x^2 = \frac{1}{2}k_b T^2 \tag{4.80}$$

where k_b is *Boltzmann's constant* and T is the *temperature* in degrees absolute, one gets:

$$PV = NkT \tag{4.81}$$

which is the ***Ideal Gas Law***[122] .

This all rather amazing and useful, and is generally covered and/or derived in a thermodynamics course, but is a bit beyond our scope for this semester. It's an excellent use of impulse, though, and the homework problem involving bouncing of a stream of beads off of the pan of a scale is intended to be "practice" for doing it then, or at least reinforcing the understanding of how pressure arises for later on in *this* course when we treat fluids.

In the meantime, the impulse approximation reduces a potentially *complicated* force of interaction during a collision to its most basic parameters – the change in momentum it causes and the (short) time over which it occurs. Life is simple, life is good. Momentum conservation (as an equation or set of equations) will yield one or more relations between the various

[121] Wikipedia: http://www.wikipedia.org/wiki/Equipartition Theorem.

[122] Wikipedia: http://www.wikipedia.org/wiki/Ideal Gas Law. The physicist version of it anyway. Chemists have the pesky habit of converting the number of molecules into the number of moles using Avogadro's number $N = 6 \times 10^{23}$ and expressing it as $PV = nRT$ instead, where $R = k_b N_A$. then using truly horrendous units such as liter-atmospheres in

momentum components of the initial and final state in a collision, and with luck and enough additional data in the problem description will enable us to solve them simultaneously for one or more unknowns. Let's see how this works.

4.4: Center of Mass Reference Frame

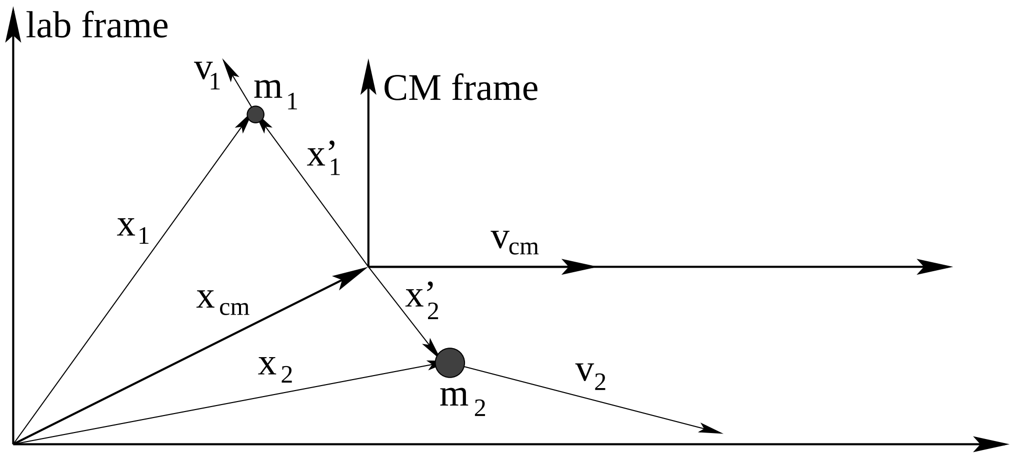


Figure 4.13: The coordinates of the "center of mass reference frame", a very useful inertial reference frame for solving collisions and understanding rigid rotation.

In the "lab frame" – the frame in which we actually live – we are often in some sense out of the picture as we try to solve physics problems, trying to make sense of the motion of flies buzzing around in a moving car as it zips by us. In the **Center of Mass Reference Frame** we are *literally* in the middle of the action, watching the flies *in* the frame of the moving car, or standing a ground zero for an impending collision. This makes it a very convenient frame for analyzing collisions, rigid rotations around an axis through the center of mass (which we'll study next week), static equilibrium (in a couple more weeks). At the end of this week, we will also derive a crucial result connecting the kinetic energy of a system of particles in the lab to the kinetic energy of the same system evaluated in the center of mass frame that will help us understand how work or mechanical energy can be transformed without loss into ***enthalpy*** (the heating of an object) during a collision or to ***rotational kinetic energy*** as an object rolls!

Recall from Week 2 the *Galilean transformation* between two ***inertial references frames*** where the primed one is moving at constant velocity $\vec{v}_{\rm frame}$ compared to the unprimed (lab) reference frame, equation 2.72.

$$\vec{x}'_i = \vec{x}_i - \vec{v}_{\rm frame} t \tag{4.82}$$

We choose our lab frame so that at time $t = 0$ the origins of the two frames are the same for simplicity. Then we take the time derivative of this equation, which connects the velocity in the lab frame to the velocity in the moving frame:

$$\vec{v}'_i = \vec{v}_i - \vec{v}_{\rm frame} \tag{4.83}$$

I always find it handy to have a simple conceptual metaphor for this last equation: The velocity of flies observed *within* a moving car equals the velocity of the flies as seen by an observer on the ground minus the velocity of the car, or equivalently the velocity seen on the ground is the velocity of the car plus the velocity of the flies measured relative to the car. That helps me get the sign in the transformation correct without having to draw pictures or do actual algebra.

Let's define the Center of Mass Frame to be the particular frame whose origin is at the center of mass of a collection of particles that have no external force acting on them, so that the total momentum of the system is ***constant*** and the velocity of the center of mass of the system is also ***constant***:

$$\vec{\boldsymbol{P}}_{\rm tot} = M_{\rm tot}\vec{\boldsymbol{v}}_{\rm cm} = \text{a constant vector} \tag{4.84}$$

or (dividing by $M_{\rm tot}$ and using the definition of the velocity of the center of mass):

$$\vec{\boldsymbol{v}}_{\rm cm} = \frac{1}{M_{\rm tot}}\sum_i m_i\vec{\boldsymbol{v}}_i = \text{a constant vector.} \tag{4.85}$$

Then the following two equations define the Galilean transformation of position and velocity coordinates from the (unprimed) lab frame into the (primed) center of mass frame:

$$\vec{\boldsymbol{x}}_i' = \vec{\boldsymbol{x}}_i - \vec{\boldsymbol{x}}_{\rm cm} = \vec{\boldsymbol{x}}_i - \vec{\boldsymbol{v}}_{\rm cm}t \tag{4.86}$$

$$\vec{\boldsymbol{v}}_i' = \vec{\boldsymbol{v}}_i - \vec{\boldsymbol{v}}_{\rm cm} \tag{4.87}$$

An enormously useful property of the center of mass reference frame follows from adding up the total momentum in the center of mass frame:

$$\begin{aligned}
\vec{\boldsymbol{P}}_{\rm tot}' &= \sum_i m_i\vec{\boldsymbol{v}}_i' = \sum_i m_i(\vec{\boldsymbol{v}}_i - \vec{\boldsymbol{v}}_{\rm cm}) \\
&= (\sum_i m_i\vec{\boldsymbol{v}}_i) - (\sum_i m_i)\vec{\boldsymbol{v}}_{\rm cm} \\
&= M_{\rm tot}\vec{\boldsymbol{v}}_{\rm cm} - M_{\rm tot}\vec{\boldsymbol{v}}_{\rm cm} = 0 \quad (!)
\end{aligned} \tag{4.88}$$

The total momentum *in* the center of mass frame is identically zero! In retrospect, this is obvious. The center of mass is at the origin, at rest, in the center of mass frame ***by definition***, so its velocity $\vec{\boldsymbol{v}}_{\rm cm}'$ is zero, and therefore it should come as no surprise that $\vec{\boldsymbol{P}}_{\rm tot}' = M_{\rm tot}\vec{\boldsymbol{v}}_{\rm cm}' = 0$.

As noted above, the center of mass frame will be very useful to us both conceptually and computationally. Our first application of the concept will be in analyzing collisions. Let's get started!

4.5: Collisions

A "collision" in physics occurs when two bodies that are more or less *not* interacting (because they are too far apart to interact) come "in range" of their mutual interaction force, strongly interact for a short time, and then separate so that they are once again too far apart to interact. We usually think of this in terms of "before" and "after" states of the system – a collision takes a pair of particles from having some known initial "free" state right before the interaction occurs to an unknown final "free" state right after the interaction occurs. A good mental model for the interaction force (as a function of time) during the collision is the impulse force sketched above that is zero at all times but the short time Δt that the two particles are in range and strongly interacting.

There are three general "types" of collision:

- Elastic
- Fully Inelastic
- Partially Inelastic

In this section, we will first indicate a single universal assumption we will make when solving scattering problems using *kinematics* (conservation laws) as opposed to *dynamics* (solving the actual equations of motion for the interaction through the collision). Next, we will briefly define each type of collision listed above. Finally, in the following sections we'll spend some time studying each type in some detail and deriving solutions where it is not too difficult.

4.5.1: Momentum Conservation in the Impulse Approximation

Most collisions that occur rapidly enough to be treated in the impulse approximation conserve momentum even if the particles are not exactly free before and after (because they are moving in a gravitational field, experiencing drag, etc). There are, of course, exceptions – cases where the collision occurs slowly and with weak forces compared to external forces, and the most important exception – collisions with objects connected to (usually much larger) "immobile" objects by a pivot or via contact with some surface.

An example of the latter is dropping one pool ball on another that is resting on a table. As the upper ball collides with the lower, the impulse the second ball experiences is communicated to the table, where it generates an impulse in the normal force there preventing that ball from moving! Momentum can hardly be conserved unless one includes the table and the entire Earth (that the table, in turn, sits on) in the calculation!

This sort of thing will *often* be the case when we treat rotational collisions in a later chapter, where disks or rods are ***pivoted*** via a connection to a large immobile object (essentially, the Earth). During the collision the collision impulse will ***often generate an impulse force at the pivot*** and cause momentum ***not*** to be conserved between the two colliding bodies.

On the other hand, in many other cases the external forces acting on the two bodies will not be "hard" constraint forces like a normal force or a pivot of some sort. Things like gravity, friction, drag, and spring forces will usually be much smaller than the impulse force and will not *change* due to the impulse itself, and hence are ignorable in the impulse approximation. For this reason, the validity of the impulse approximation will be our ***default assumption*** in the collisions we treat in this chapter, and hence we will assume that ***all collisions conserve total momentum*** through the collision unless we can *see* a pivot or normal force that will exert a counter-impulse of some sort.

To summarize, whether or not any "soft" external force is acting during the collision, we will make the impulse approximation and assume that the total vector momentum of the colliding particles right *before* the collision will equal the total vector momentum of the colliding particles right *after* the collision.

Because momentum is a three-dimensional vector, this yields one to three (relevant) independent equations that *constrain* the solution, depending on the number of dimensions in which the collision occurs.

4.5.2: Elastic Collisions

By definition, an ***elastic collision*** is one that ***also*** conserves ***total kinetic energy*** so that the total scalar kinetic energy of the colliding particles before the collision must equal the total kinetic energy after the collision. This is an additional independent equation that the solution must satisfy.

It is assumed that all other contributions to the total mechanical energy (for example, gravitational potential energy) are identical before and after if not just zero, again this *is* the impulse approximation that states that all of these forces are negligible compared to the collision force over the time Δt. However, two of your homework problems will treat exceptions by explicitly *giving* you a conservative, "slow" interaction force (gravity and an inclined plane slope, and a spring) that *mediates* the "collision". You can use these as mental models for what really happens in elastic collisions on a much faster and more violent time frame.

4.5.3: Fully Inelastic Collisions

For *inelastic* collisions, we will assume that the two particles form a single "particle" as a final state with the same total momentum as the system had before the collision. In these collisions, kinetic energy is *always lost*. Since energy itself is technically conserved, we can ask ourselves: Where did it go? The answer is: Into *heat*[123]*!* The final section in this chapter "discovers" that we have completely neglected both organized and disorganized/internal energy by treating extended objects (which are really "systems") as if they are particles.

One important characteristic of fully inelastic collisions, and the property that distinguishes them from partially inelastic collisions, is that the energy lost to heat in a fully inelastic collision is the *maximum* energy that *can* be lost in a momentum-conserving collision, as will be proven and discussed below.

Inelastic collisions are much easier to solve than elastic (or partially inelastic) ones, because there are fewer degrees of freedom in the final state (only one velocity, not two). As we

[123]Or more properly, into Enthalpy, which is *microscopic* mechanical energy distributed among the atoms and molecules that make up an object. Also into things like sound, light, the energy carried away by flying debris if any.

count this up later, we will see that inelastic collisions are *fully solvable* using kinematics alone, independent of the details of the mediating force and without additional information.

4.5.4: Partially Inelastic Collisions

As suggested by their name, a partially inelastic collision is one where *some* kinetic energy is lost in the collision (so it isn't elastic) but not the maximum amount. The particles do not stick together, so there are in general two velocities that must be solved for in the "after" picture, just as there are for elastic collisions. In general, since *any* energy from zero (elastic) to some maximum amount (fully inelastic) can be lost during the collision, you will have to be given more information about the problem (such as the velocity of one of the particles after the collision) in order to be able to solve for the remaining information and answer questions.

4.5.5: Dimension of Scattering and Sufficient Information

Given an actual force law describing a collision, one can in principle always solve the dynamical differential equations that result from applying Newton's Second Law to all of the masses and find their final velocities from their initial conditions and a knowledge of the interaction force(s). However, the solution of collisions involving all but the *simplest* interaction forces is beyond the scope of this course (and is usually quite difficult).

The reason for defining the collision types above is because they all represent *kinematic* (math with units) constraints that are true independent of the details of the interaction force beyond it being either conservative (elastic) or non-conservative (fully or partially inelastic). In some cases the kinematic conditions alone are sufficient to solve the entire scattering problem! In others, however, one cannot obtain a final answer without knowing the details of the scattering force as well as the initial conditions, or without knowing *some* of the details of the final state.

To understand this, consider only elastic collisions. If the collision occurs in three dimensions, one has four equations from the kinematic relations – three independent momentum conservation equations (one for each component) plus one equation representing kinetic energy conservation. However, the outgoing particle velocities have *six* numbers in them – three components each. There simply aren't enough kinematic constraints to be able to predict the final state from the initial state without knowing the interaction.

Many collisions occur in two dimensions – think about the game of pool, for example, where the cue ball "elastically" strikes the eight ball. In this case one has two momentum conservation equations and one energy conservation equation, but one needs to solve for the four components of two final velocities in two dimensions. Again we either need to know *something* about the velocity of *one* of the two outgoing particles – say, its x-component – or we cannot solve for the remaining components without a knowledge of the interaction.

Of course in the game of pool[124] we *do* know something very important about the interaction. It is a force that is exerted directly along the line connecting the centers of the balls at the instant they strike one another! This is just enough information for us to be able to mentally

[124] Or "billiards".

predict that the eight ball will go into the corner pocket if it begins at rest and is struck by the cue ball on the line from that pocket back through the center of the eight ball. This in turn is sufficient to predict the trajectory of the cue ball as well.

Two dimensional elastic collisions are thus *almost* solvable from the kinematics. This makes them too difficult for students who are unlikely to spend much time analyzing actual collisions (although it is worth it to look them over in the specific context of a good example, one that many students have direct experience with, such as the game of pool/billiards). Physics majors should spend some time here to prepare for more difficult problems later, but life science students can probably skip this without any great harm.

One dimensional elastic collisions, on the other hand, have one momentum conservation equation and one energy conservation equation to use to solve for two unknown final velocities. The number of independent equations and unknowns match! We can thus solve one dimensional elastic collision problems *without knowing the details of the collision force* from the kinematics alone.

Things are much simpler for fully *in*elastic collisions. Although one only has one, two, or three momentum conservation equations, this precisely matches the number of components in the final velocity of the two masses moving together as one after they have stuck together! The final velocity is thus fully determined from the initial velocities (and hence total momentum) of the colliding objects. Fully inelastic collisions are thus the *easiest* collision problems to solve in any number of dimensions.

Partially inelastic collisions in any number of dimensions are the most difficult to solve or least determined (from the kinematic point of view). There one *loses the energy conservation equation* – one cannot even solve the *one dimensional* partially inelastic collision problem without either being given some additional information about the final state – typically the final velocity of one of the two particles so that the other can be found from momentum conservation – or solving the dynamical equations of motion, which is generally even more difficult.

This explains why this textbook focuses on only four relatively simple characteristic collision problems. We first study elastic collisions in one dimension, solving them in two slightly different ways that provide different insights into how the physics works out. I then talk briefly about elastic collisions in two dimensions in an "elective" section that can safely be omitted by non-physics majors (but is quite readable, I hope). We then cover inelastic collisions in one and two dimensions, concentrating on the fully solvable case (fully inelastic) but providing a simple example or two of partially inelastic collisions as well.

4.6: 1-D Elastic Collisions

In figure 4.14 above, we see a typical one-dimensional collision between two masses, m_1 and m_2. m_1 has a speed in the x-direction $v_{1i} > 0$ and m_2 has a speed $v_{2i} < 0$, but our solution should not only handle the specific picture above, it should also handle the (common) case where m_2 is initially at rest ($v_{2i} = 0$) or even the case where m_2 is moving to the right, but more slowly than m_1 so that m_1 overtakes it and collides with it, $v_{1i} > v_{2i} > 0$. Finally, there is nothing special about the labels "1" and "2" – our answer should be symmetric (still work if we label the mass on the left 2 and the mass on the right 1).

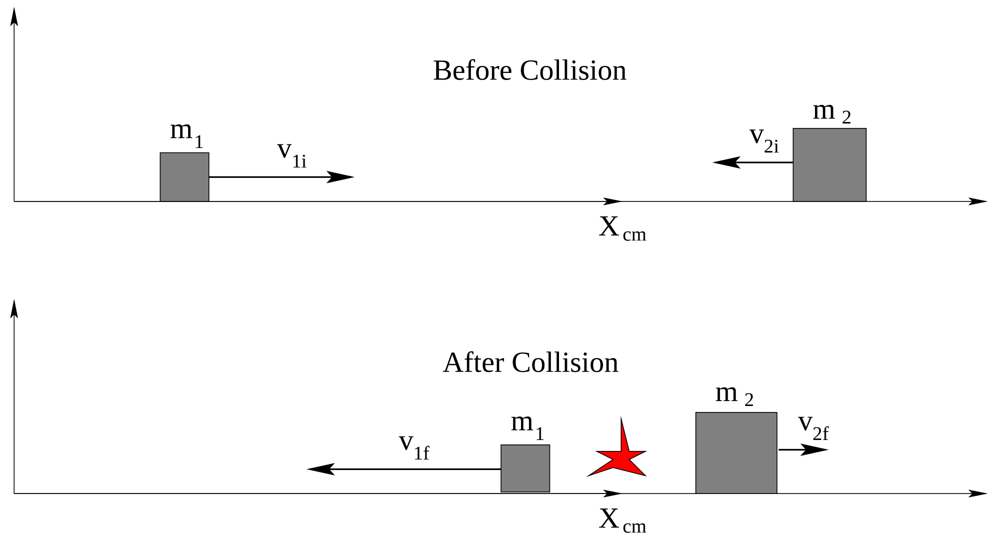


Figure 4.14: Before and after snapshots of an elastic collision in one dimension, illustrating the important quantities.

We seek final velocities that satisfy the two conditions that define an elastic collision.
Momentum Conservation:

$$\begin{aligned} \vec{\boldsymbol{p}}_{1i} + \vec{\boldsymbol{p}}_{2i} &= \vec{\boldsymbol{p}}_{1f} + \vec{\boldsymbol{p}}_{2f} \\ m_1 v_{1i}\hat{\boldsymbol{x}} + m_2 v_{2i}\hat{\boldsymbol{x}} &= m_1 v_{1f}\hat{\boldsymbol{x}} + m_2 v_{2f}\hat{\boldsymbol{x}} \quad (4.89) \\ m_1 v_{1i} + m_2 v_{2i} &= m_1 v_{1f} + m_2 v_{2f} \quad (x\text{-direction only}) \quad (4.90) \end{aligned}$$

Kinetic Energy Conservation:

$$\begin{aligned} E_{k1i} + E_{k2i} &= E_{k1f} + E_{k2f} \\ \frac{1}{2}m_1 v_{1i}^2 + \frac{1}{2}m_2 v_{2i}^2 &= \frac{1}{2}m_1 v_{1f}^2 + \frac{1}{2}m_2 v_{2f}^2 \quad (4.91) \end{aligned}$$

Note well that although our figure shows m_2 moving to the left, we expressed momentum conservation without an assumed minus sign! Our solution has to be able to handle ***both*** positive ***and*** negative velocities for either mass, so we will assume them to be positive in our equations and simply use a negative value for e.g. v_{2i} if it happens to be moving to the left in an actual problem we are trying to solve.

The big question now is: Assuming we know m_1, m_2, v_{1i} and v_{2i}, can we find v_{1f} and v_{2f}, even though we have not specified any of the details of the interaction between the two masses during the collision? This is not a trivial question! In three dimensions, the answer might well be *no*, not without more information. In one dimension, however, we have two independent equations and two unknowns, and it turns out that these two conditions alone suffice to determine the final velocities.

To get this solution, we must solve the two conservation equations above simultaneously. There are three ways to proceed.

One is to use simple substitution – manipulate the momentum equation to solve for (say) v_{2f} in terms of v_{1f} and the givens, substitute it into the energy equation, and then brute force solve the energy equation for v_{1f} and back substitute to get v_{2f}. This involves solving an annoying quadratic (and a horrendous amount of intermediate algebra) and in the end, gives us no insight at all into the conceptual "physics" of the solution. We will therefore avoid it, although if one has the patience and care to work through it it will give one the right answer.

The second approach is basically a much better/smarter (but perhaps less obvious) algebraic solution, and gives us at least *one* important insight. We will treat it – the "relative velocity" approach – first in the subsections below.

The third is the most informative, and (in my opinion) the ***simplest***, of the three solutions – once one has mastered the *concept* of the center of mass reference frame outlined above. This "center of mass frame" approach (where the collision occurs right in front of your eyes, as it were) is the one I suggest that all students learn, because it can be reduced to four very simple steps and because it yields by far the most *conceptual* understanding of the scattering process.

4.6.1: The Relative Velocity Approach

As I noted above, using direct substitution openly invites madness and frustration for all but the most skilled young algebraists. Instead of using substitution, then, let's rearrange the energy conservation equation and momentum conservation equations to get all of the terms with a common mass on the same side of the equals signs and do a bit of simple manipulation of the energy equation as well:

$$\begin{aligned} m_1 v_{1i}^2 - m_1 v_{1f}^2 &= m_2 v_{2f}^2 - m_2 v_{2i}^2 \\ m_1(v_{1i}^2 - v_{1f}^2) &= m_2(v_{2f}^2 - v_{2i}^2) \\ m_1(v_{1i} - v_{1f})(v_{1i} + v_{1f}) &= m_2(v_{2f} - v_{2i})(v_{2i} + v_{2f}) \end{aligned} \tag{4.92}$$

(from energy conservation) and

$$m_1(v_{1i} - v_{1f}) = m_2(v_{2f} - v_{2i}) \tag{4.93}$$

(from momentum conservation).

When we divide the first of these by the second (subject to the condition that $v_{1i} \neq v_{1f}$ and $v_{2i} \neq v_{2f}$ to avoid dividing by zero, a condition that incidentally guarantees that *a collision occurs* as one possible solution to the kinematic equations alone is *always* for the final velocities to equal the initial velocities, meaning that no collision occured), we get:

$$(v_{1i} + v_{1f}) = (v_{2i} + v_{2f}) \tag{4.94}$$

or (rearranging):

$$(v_{2f} - v_{1f}) = -(v_{2i} - v_{1i}) \tag{4.95}$$

This final equation can be interpreted as follows in English: ***The relative velocity of recession after a collision equals (minus) the relative velocity of approach before a collision.*** This is an important *conceptual* property of elastic collisions.

Although it isn't obvious, this equation is independent from the momentum conservation equation and can be used with it to solve for v_{1f} and v_{2f}, e.g. –

$$\begin{aligned} v_{2f} &= v_{1f} - (v_{2i} - v_{1i}) \\ m_1 v_{1i} + m_2 v_{2i} &= m_1 v_{1f} + m_2(v_{1f} - (v_{2i} - v_{1i})) \\ (m_1 + m_2) v_{1f} &= (m_1 - m_2) v_{1i} + 2 m_2 v_{2i} \end{aligned} \tag{4.96}$$

Instead of just solving for v_{1f} and either backsubstituting or invoking symmetry to find v_{2f} we now work a bit of algebra magic that you won't see the point of until the end. Specifically, let's add zero to this equation by adding and subtracting $m_1 v_{1i}$:

$$\begin{aligned} (m_1 + m_2) v_{1f} &= (m_1 - m_2) v_{1i} + 2 m_2 v_{2i} + (m_1 v_{1i} - m_1 v_{1i}) \\ &= -(m_1 + m_2) v_{1i} + 2\,(m_2 v_{2i} + m_1 v_{1i}) \end{aligned} \tag{4.97}$$

(check this on your own). Finally, we divide through by $m_1 + m_2$ and get:

$$v_{1f} = -v_{1i} + 2\frac{m_1 v_{1i} + m_2 v_{2i}}{m_1 + m_2} \tag{4.98}$$

The last term is just two times the total initial momentum divided by the total mass, which we should *recognize* to be able to write:

$$v_{1f} = -v_{1i} + 2 v_{\mathrm{cm}} \tag{4.99}$$

There is nothing special about the labels "1" and "2", so the solution for mass 2 must be *identical*:

$$v_{2f} = -v_{2i} + 2 v_{\mathrm{cm}} \tag{4.100}$$

although you can also obtain this directly by backsubstituting v_{1f} into equation 4.95.

This solution looks simple enough and isn't horribly difficult to *memorize*, but the derivation is difficult to *understand* and hence *learn*. Why do we perform the steps above, or rather, why should we have known to try those steps? The best answer is because they end up working out pretty well, a lot better than brute force substitutions (the obvious thing to try), which isn't very helpful. We'd prefer a good *reason*, one linked to our eventual conceptual understanding of the scattering process, and while equation 4.95 had a whiff of concept and depth and ability to be really learned in it (justifying the work required to obtain the result) the "magical" appearance of v_{cm} in the final answer in a very simple and symmetric way is quite mysterious (and only occurs after performing some adding-zero-in-just-the-right-form dark magic from the book of algebraic arts).

To **understand** the collision and why this in particular is the answer, it is easiest to **put everything into the *center of mass (CM) reference frame*, evaluate the collision, and then put the results back into the lab frame!** This (as we will see) naturally leads to the same result, but in a way we can *easily understand* and that gives us valuable practice in frame transformations besides!

4.6.2: 1D Elastic Collision in the Center of Mass Frame

Here is a bone-simple recipe for solving the 1D elastic collision problem in the center of mass frame.

a) Transform the problem (initial velocities) into the center of mass frame.

b) Solve the problem. The "solution" in the center of mass frame is (as we will see) ***trivial***: Reverse the center of mass velocities.

c) Transform the answer back into the lab/original frame.

Suppose as before we have two masses, m_1 and m_2, approaching each other with velocities v_{1i} and v_{2i}, respectively. We start by evaluating the velocity ***of*** the CM frame:

$$v_{\rm cm} = \frac{m_1 v_{1i} + m_2 v_{2i}}{m_1 + m_2} \tag{4.101}$$

and then transform the initial velocities ***into*** the CM frame:

$$\begin{aligned} v'_{1i} &= v_{1i} - v_{\rm cm} \qquad (4.102)\\ v'_{2i} &= v_{2i} - v_{\rm cm} \qquad (4.103)\end{aligned}$$

We know that momentum must be conserved in *any* inertial coordinate frame (in the impact approximation). In the CM frame, of course, the total momentum is *zero* so that the momentum conservation equation becomes:

$$\begin{aligned} m_1 v'_{1i} + m_2 v'_{2i} &= m_1 v'_{1f} + m_2 v'_{2f} \qquad (4.104)\\ p'_{1i} + p'_{2i} &= p'_{1f} + p'_{2f} = 0 \qquad (4.105)\end{aligned}$$

Thus $p'_i = p'_{1i} = -p'_{2i}$ and $p'_f = p'_{1f} = -p'_{2f}$. The energy conservation equation (in terms of the p's) becomes:

$$\begin{aligned} \frac{p'^2_i}{2m_1} + \frac{p'^2_i}{2m_2} &= \frac{p'^2_f}{2m_1} + \frac{p'^2_f}{2m_2} \quad \text{or}\\ p'^2_i\left(\frac{1}{2m_1} + \frac{1}{2m_2}\right) &= p'^2_f\left(\frac{1}{2m_1} + \frac{1}{2m_2}\right) \quad \text{so that}\\ p'^2_i &= p'^2_f \qquad (4.106)\end{aligned}$$

Taking the square root of both sides (and recalling that p'_i refers equally well to mass 1 or 2):

$$\begin{aligned} p'_{1f} &= \pm p'_{1i} \qquad (4.107)\\ p'_{2f} &= \pm p'_{2i} \qquad (4.108)\end{aligned}$$

The + sign rather obviously satisfies the two conservation equations. The two particles keep on going at their original speed and with their original energy! This is, actually, a perfectly good solution to the scattering problem and could be true even if the particles "hit" each other. The more interesting case (and the one that is appropriate for "hard" particles that cannot interpenetrate) is for the particles to *bounce apart* in the center of mass frame after the collision. We therefore choose the minus sign in this result:

$$\begin{aligned} p'_{1f} = m_1 v'_{1f} &= -m_1 v'_{1i} = -p'_{1i} \qquad (4.109)\\ p'_{2f} = m_2 v'_{2f} &= -m_2 v'_{2i} = -p'_{2i} \qquad (4.110)\end{aligned}$$

Since the masses are the same before and after we can divide them out of each equation and obtain the solution to the elastic scattering problem *in* the CM frame as:

$$v'_{1f} = -v'_{1i} \tag{4.111}$$
$$v'_{2f} = -v'_{2i} \tag{4.112}$$

or ***the velocities of*** m_1 ***and*** m_2 ***reverse*** in the CM frame.

This actually makes *sense*. It guarantees that if the momentum was zero before it is still zero, and since the *speed* of the particles is unchanged (only the direction of their velocity in this frame changes) the total kinetic energy is similarly unchanged.

Finally, it is trivial to put the these solutions back into the lab frame by *adding* $v_{\rm cm}$ to them:

$$\begin{aligned} v_{1f} &= v'_{1f} + v_{\rm cm} \\ &= -v'_{1i} + v_{\rm cm} \\ &= -(v_{1i} - v_{\rm cm}) + v_{\rm cm} \quad \text{or} \\ v_{1f} &= -v_{1i} + 2v_{\rm cm} \quad \text{and similarly} \qquad (4.113) \\ v_{2f} &= -v_{2i} + 2v_{\rm cm} \qquad (4.114) \end{aligned}$$

These are the exact same solutions we got in the first example/derivation above, but now they have considerably more *meaning*. The "solution" to the elastic collision problem in the CM frame is that *the velocities reverse* (which of course makes the relative velocity of approach be the negative of the relative velocity of recession, by the way). We can see that this is the solution in the center of mass frame in one dimension *without* doing the formal algebra above, it *makes sense!*

That's it then: to solve the one dimensional elastic collision problem all one has to do is transform the initial velocities into the CM frame, reverse them, and transform them back. Nothing to it.

Note that (however it is derived) these solutions are *completely symmetric* – we obviously don't care *which* of the two particles is *labelled* "1" or "2", so the answer should have exactly the same *form* for both. Our derived answers clearly have that property. In the end, we only need ***one equation*** (plus our ability to evaluate the velocity of the center of mass):

$$v_f = -v_i + 2v_{\rm cm} \tag{4.115}$$

valid for *either particle*.

If you are a physics major, you should be prepared to derive this result one of the various ways it can be derived (I'd strongly suggest the last way, using the CM frame). If you are e.g. a life science major or engineer, you should derive this result for yourself *at least once, at least one of the ways* (again, I'd suggest that last one) but then you are also welcome to memorize/learn the resulting *solution* well enough to use it.

Note well! If you remember the three steps needed for the center of mass frame derivation, even if you *forget* the actual solution on a quiz or a test – which is probably quite likely as I have little confidence in memorization as a learning tool for mountains of complicated material – you have a *prayer* of being able to rederive it on a test.

4.6.3: The "BB/bb" or "Pool Ball" Limits

In collision problems in general, it is worthwhile thinking about the "ball bearing and bowling ball (BB) limits"[125]. In the context of elastic 1D collision problems, these are basically the asymptotic results one obtains when one hits a stationary bowling ball (large mass, BB) with rapidly travelling ball bearing (small mass, bb).

This should be something you already know the answer to from experience and intuition. We all know that if you shoot a bb gun at a bowling ball so that it collides elastically, it will bounce back off of it (almost) as fast as it comes in and the bowloing ball will hardly recoil[126]. Given that $v_{\rm cm}$ in this case is more or less equal to v_{BB}, that is, $v_{\rm cm} \approx 0$ (just a bit greater), note that this is *exactly* what the solution predicts.

What happens if you throw a bowling ball at a stationary bb? Well, we know perfectly well that the BB in this case will just continue barrelling along at more or less $v_c m$ (still roughly equal to the velocity of the more massive bowling ball) – ditto, when your car hits a bug with the windshield, it doesn't significantly slow down. The bb (or the bug) on the other hand, *bounces forward off of the BB* (or the windshield)!

In fact, according to our results above, it will bounce off the BB and recoil forward at approximately *twice the speed* of the BB. Note well that both of these results preserve the idea derived above that the relative velocity of approach equals the relative velocity of recession, and you can transform from one to the other by just changing your frame of reference to ride along with BB or bb – two different ways of looking at the same collision.

Finally, there is the "pool ball limit" – the elastic collision of roughly equal masses. When the cue ball strikes another ball head on (with no English), then as pool players well know the cue ball stops (nearly) dead and the other ball continues on at the original speed of the cue ball. This, too, is exactly what the equations/solutions above predict, since in this case $v_{\rm cm} = v_{1i}/2$.

Our solutions thus agree with our experience and intuition in *both* the limits where one mass is much larger than the other *and* when they are both roughly the same size. One has to expect that they are probably valid everywhere. Any answer you derive (such as this one) ultimately has to pass the test of common-sense agreement with your everyday experience. This one seems to, however difficult the derivation was, it appears to be correct!

As you can probably guess from the extended discussion above, pool is a good example of a game of "approximately elastic collisions" because the hard balls used in the game have a very elastic coefficient of restitution, another way of saying that the surfaces of the balls behave like very small, very hard springs and store and re-release the kinetic energy of the collision from a conservative impulse type force.

However, it also opens up the question: What happens if the collision between two balls is *not* along a line? Well, then we have to take into account momentum conservation in *two* dimensions. So alas, my fellow human students, we are all going to have to bite the bullet and at least think a *bit* about collisions in more than one dimension.

[125] Also known as the "windshield and bug limits"...

[126] ...and you'll put your eye out – kids, do *not* try this at home!

4.7: Elastic Collisions in 2-3 Dimensions

As we can see, elastic collisions in one dimension are "good" because we can completely solve them using only kinematics – we don't care about the *details* of the interaction between the colliding entities; we can find the final state from the initial state for all possible elastic forces and the only differences that will depend on the forces will be things like how *long* it takes for the collision to occur.

In 2+ dimensions we at the very least have to work much harder to solve the problem. We will no longer be able to use nothing but vector momentum conservation and energy conservation to solve the problem independent of most of the details of the interaction. In two dimensions we have to solve for four outgoing components of velocity (or momentum), but we only have conservation equations for two components of momentum and kinetic energy. Three equations, four unknowns means that the problem is indeterminate *unless* we are told at least one more thing about the final state, such as one of the components of the velocity or momentum of one of the outgoing masses. In three dimensions it is even worse – we must solve for six outgoing components of velocity/momentum but have only four conservation equations (three momentum, one energy) and need at least two additional pieces of information. Kinematics alone is simply insufficient to solve the scattering problem – need to know the details of the potential/force of interaction and solve the equations of motion for the scattering in order to predict the final/outgoing state from a knowledge of the initial/incoming state.

The dependence of the outoing scattering on the interaction is good and bad. The good thing is that we can *learn* things about the interaction from the results of a collision experiment (in one dimension, note well, our answers didn't depend on the interaction force so we learn nothing at all about that force aside from the fact that it is elastic from scattering data). The bad is that for the most part the algebra and *calculus* involved in solving multidimensional collisions is well beyond the scope of this course. Physics majors, and perhaps a few other select individuals in other majors or professions, will have to sweat blood *later* to work all this out for a tiny handful of interaction potentials where the problem is analytically solvable, but *not yet!*

Still, there are a few things that *are* within the scope of the course, at least for majors. These involve learning a bit about how to set up a good coordinate frame for the scattering, and how to treat "hard sphere" elastic collisions which turn out to be *two* dimensional, and hence solvable from kinematics plus a single assumption about recoil direction in at least some simple cases. Let's look at scattering in two dimensions in the case where the target particle is at rest and the outgoing particles lie (necessarily) in a plane.

We expect both energy and momentum to be conserved in any elastic collision. This gives us the following set of equations:

$$p_{0x} = p_{1x} + p_{2x} \tag{4.116}$$

$$p_{0y} = p_{1y} + p_{2y} \tag{4.117}$$

$$p_{0z} = p_{1z} + p_{2z} \tag{4.118}$$

(for momentum conservation) and

$$\frac{p_0^2}{2m_1} = E_0 = E_1 + E_2 = \frac{p_1^2}{2m_1} + \frac{p_2^2}{m_2} \tag{4.119}$$

for kinetic energy conservation.

We have four equations, and four unknowns, so we *might* hope to be able to solve it quite generally. However, we don't *really* have that many equations – if we assume that the scattering plane is the $x-y$ plane, then necessarily $p_{0z} = p_{1z} = p_{2z} = 0$ and this equation tells us nothing useful. We need more information in order to be able to solve the problem.

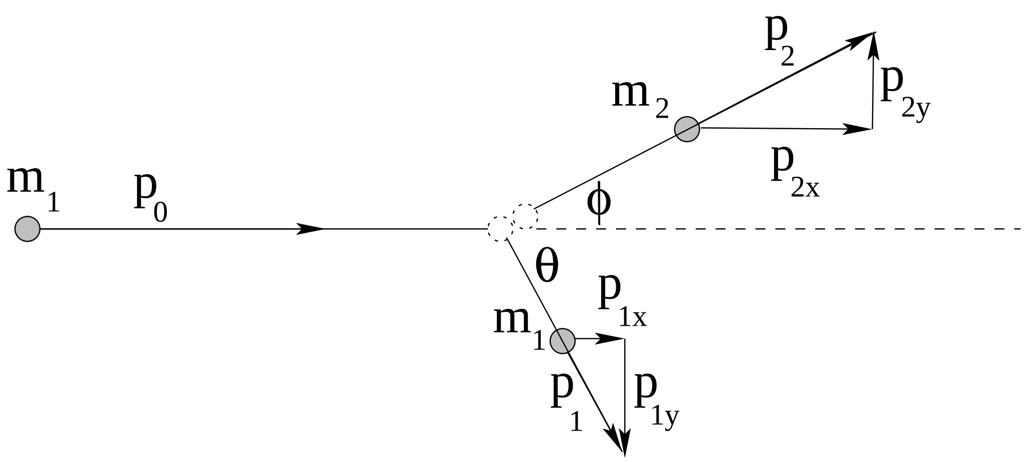


Figure 4.15: The geometry for an elastic collision in a two-dimensional plane.

Let's see what we *can* tell in this case. Examine figure 4.15. Note that we have introduced two angles: θ and ϕ for the incident and target particle's outgoing angle with respect to the incident direction. Using them and setting $p_{0y} = p_{0z} = 0$ (and assuming that the target is at rest initially and has no momentum at all initially) we get:

$$p_{0x} = p_{1x} + p_{2x} = p_1 \cos(\theta) + p_2 \cos(\phi) \tag{4.120}$$

$$p_{0y} = p_{1y} + p_{2y} = 0 = -p_1 \sin(\theta) + p_2 \sin(\phi) \tag{4.121}$$

In other words, the momentum in the x-direction is conserved, and the momentum in the y-direction (after the collision) cancels. The latter is a powerful relation – if we know the y-momentum of one of the outgoing particles, we know the other. If we know the magnitudes/energies of both, we know an important relation between their angles.

This, however, puts us no closer to being able to solve the general problem (although it does help with a special case that is on your homework). To make real progress, it is necessarily to once again change to the center of mass reference frame by subtracting $\vec{v}_{\rm cm}$ from the velocity of both particles. We can easily do this:

$$\vec{p}\,'_{i1} = m_1(\vec{v}_0 - \vec{v}_{\rm cm}) = m_1 u_1 \tag{4.122}$$

$$\vec{p}\,'_{i2} = -m_2\vec{v}_{\rm cm} = m_2 u_2 \tag{4.123}$$

so that $\vec{p}\,'_{i1} + \vec{p}\,'_{i2} = \vec{p}\,'_{\rm tot} = 0$ in the center of mass frame as usual. The initial energy in the center of mass frame is just:

$$E_i = \frac{p_{i1}^{2'}}{2m_1} + \frac{p_{i2}^{2'}}{2m_2} \tag{4.124}$$

Since $p'_{i1} = p'_{i2} = p'_i$ (the magnitudes are equal) we can simplify this a bit further:

$$E_i = \frac{p_i^{2'}}{2m_1} + \frac{p_i^{2'}}{2m_2} = \frac{p_i^{2'}}{2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right) = \frac{p_i^{2'}}{2}\left(\frac{m_1 + m_2}{m_1 m_2}\right) \tag{4.125}$$

After the collision, we can see by inspection of

$$E_f = \frac{p_f^{2'}}{2m_1} + \frac{p_f^{2'}}{2m_2} = \frac{p_f^{2'}}{2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right) = \frac{p_f^{2'}}{2}\left(\frac{m_1 + m_2}{m_1 m_2}\right) = E_i \tag{4.126}$$

that $p'_{f1} = p'_{f2} = p'_f = p'_i$ will cause energy to be conserved, just as it was for a 1 dimensional collision. All that can change, then, is the *direction* of the incident momentum in the center of mass frame. In addition, since the total momentum in the center of mass frame is by definition zero before and after the collision, if we know the *direction* of either particle after the collision in the center of mass frame, the other is the opposite:

$$\vec{p}'_{f1} = -\vec{p}'_{f2} \tag{4.127}$$

We have then "solved" the collision as much as it can be solved. We cannot *uniquely* predict the direction of the final momentum of either particle in the center of mass (or any other) frame without knowing more about the interaction and e.g. the incident impact parameter. We can predict the magnitude of the outgoing momenta, and if we know the outgoing direction alone of either particle we can find everything – the magnitude and direction of the other particle's momentum and the magnitude of the momentum of the particle whose angle we measured.

As you can see, this is all pretty difficult, so we'll leave it at this point as a partially solved problem, ready to be tackled again for specific interactions or collision models in a future course.

4.8: Inelastic Collisions

A fully inelastic collision is where two particles collide and *stick together*. As always, momentum is conserved in the impact approximation, but now *kinetic energy is not!* In fact, we will see that macroscopic kinetic energy is always *lost* in an inelastic collision, either to heat or to some sort of mechanism that traps and reversibly stores the energy.

These collisions are much easier to understand and analyze than elastic collisions. That is because there are *fewer degrees of freedom* in an inelastic collision – we can easily solve them even in 2 or 3 dimensions. The whole solution is developed from

$$\vec{p}_{i,\mathrm{ntot}} = m_1\vec{v}_{1i} + m_2\vec{v}_{2i} = (m_1 + m_2)\vec{v}_f = (m_1 + m_2)\vec{v}_{\mathrm{cm}} = \vec{p}_{f,\mathrm{tot}} \tag{4.128}$$

In other words, in a fully inelastic collision, the velocity of the outgoing combined particle is the velocity of the *center of mass* of the system, which we can easily compute from a knowledge of the initial momenta or velocities and masses. Of course! How obvious! How easy!

From this relation you can easily find $\vec{v}_f$ in any number of dimensions, and answer many related questions. The collision is "solved". However, there are a number of different kinds of *problems* one can solve given this basic solution – things that more or less tag additional physics problems on to the end of this initial one and use its result as their *starting* point, so you have to solve two or more subproblems in one long problem, one of which is the "inelastic collision". This is best illustrated in some archetypical examples.

Example 4.8.1: One-dimensional Fully Inelastic Collision (only)

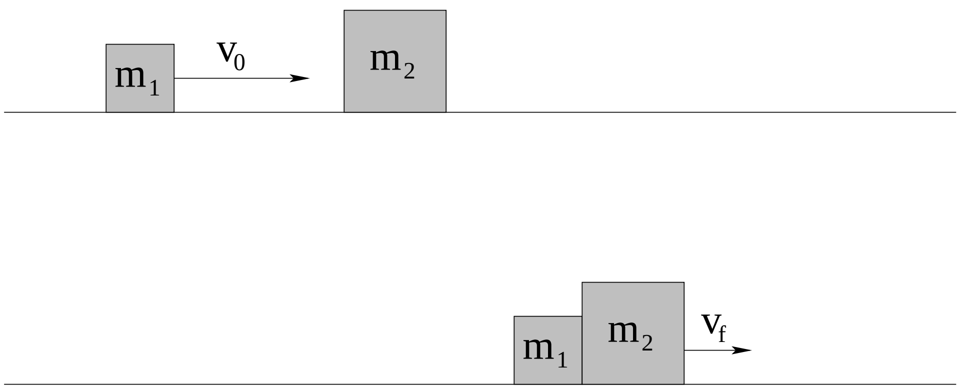


Figure 4.16: Two blocks of mass m_1 and m_2 collide and stick together on a frictionless table.

In figure 4.16 above, a block m_1 is sliding across a frictionless table at speed v_0 to strike a second block m_2 initially at rest, whereupon they stick together and move together as one thereafter at some final speed v_f.

Before, after, and during the collision, gravity acts but is opposed by a normal force. There is no friction or drag force doing any work. The only forces in play are the *internal* forces mediating the collision and making the blocks stick together. We therefore know that **momentum is conserved** in this problem independent of the features of that internal interaction. Even if friction or drag forces *did* act, as long as the collision took place "instantly" in the impact approximation, momentum would still be conserved from immediately before to immediately after the collision, when the impulse Δp of the collision force would be much, much larger than any change in momentum due to the drag over the same small time Δt.

Thus:

$$p_i = m_1 v_0 = (m_1 + m_2) v_f = p_f \tag{4.129}$$

or

$$v_f = \frac{m v_0}{(m_1 + m_2)} (\quad = v_{\rm cm}) \tag{4.130}$$

A traditional question that accompanies this is: How much kinetic energy was lost in the collision? We can answer this by simply figuring it out.

$$\begin{aligned}
\Delta K &= K_f - K_i = \frac{p_f^2}{2(m_1+m_2)} - \frac{p_i^2}{2m_1} \\
&= \frac{p_i^2}{2}\left(\frac{1}{(m_1+m_2)} - \frac{1}{m_1}\right) \\
&= \frac{p_i^2}{2}\left(\frac{m_1-(m_1+m_2)}{m_1(m_1+m_2)}\right) \\
&= -\frac{p_i^2}{2m_1}\left(\frac{m_2}{(m_1+m_2)}\right) \\
&= -\left(\frac{m_2}{(m_1+m_2)}\right)K_i
\end{aligned} \tag{4.131}$$

where we have expressed the result as a *fraction of the initial kinetic energy!*

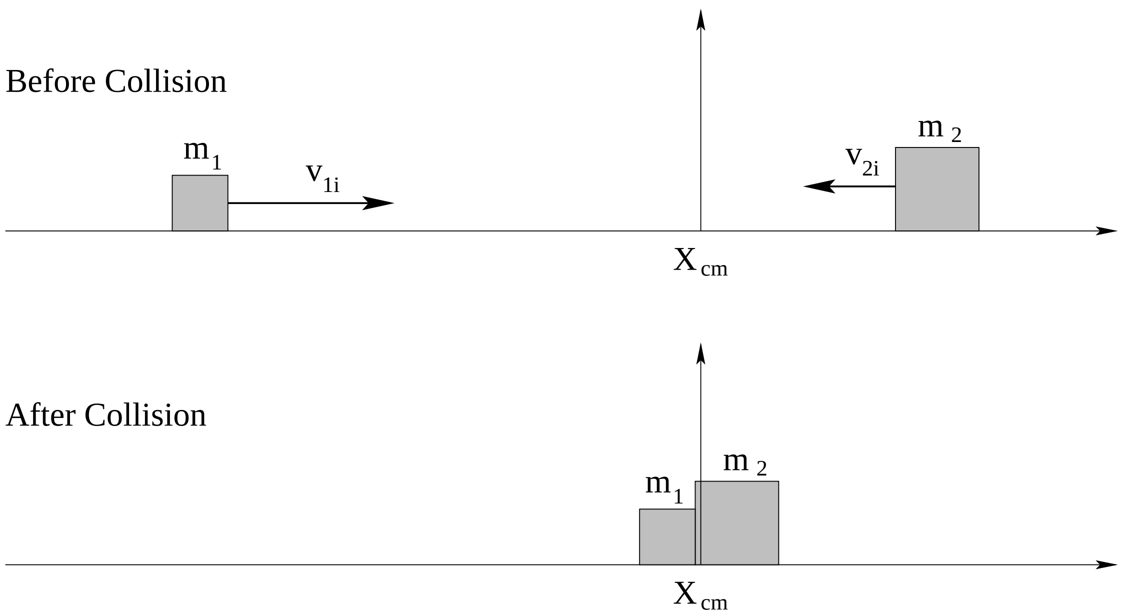


Figure 4.17: Two blocks collide and stick together on a frictionless table – in the center of mass frame. After the collision they are both at rest at the center of mass and *all of the kinetic energy they had before the collision in this frame* is lost.

There is a different way to think about the collision and energy loss. In figure 4.17 you see the same collision portrayed in the CM frame. In this frame, the two particles *always come together and stick* to remain, at rest, at the center of mass after the collision. ***All of the kinetic energy in the CM frame is lost in the collision!*** That's exactly the amount we just computed, but I'm leaving the proof of that as an exercise for you.

Note well the BB limits: For a light bb (m_1) striking a massive BB (m_2), nearly *all* the energy is lost. This sort of collision between an asteroid (bb) and the earth (BB) caused at least one of the mass extinction events, the one that ended the Cretaceous and gave mammals the leg up that they needed in a world dominated (to that point) by dinosaurs. For a massive BB (m_1) stricking a light bb (m_2) very little of the energy of the massive object is lost. Your truck hardly slows when it smushes a bug "inelastically" against the windshield. In the equal billiard ball bb collision ($m_1 = m_2$), exactly one half of the initial kinetic energy is lost.

A similar collision in 2D is given for your homework, where a truck and a car inelastically collide and then slide down the road together. In this problem friction works, but *not during the collision!* Only after the "instant" (impact approximation) collision do we start to worry about the effect of friction.

Example 4.8.2: Ballistic Pendulum

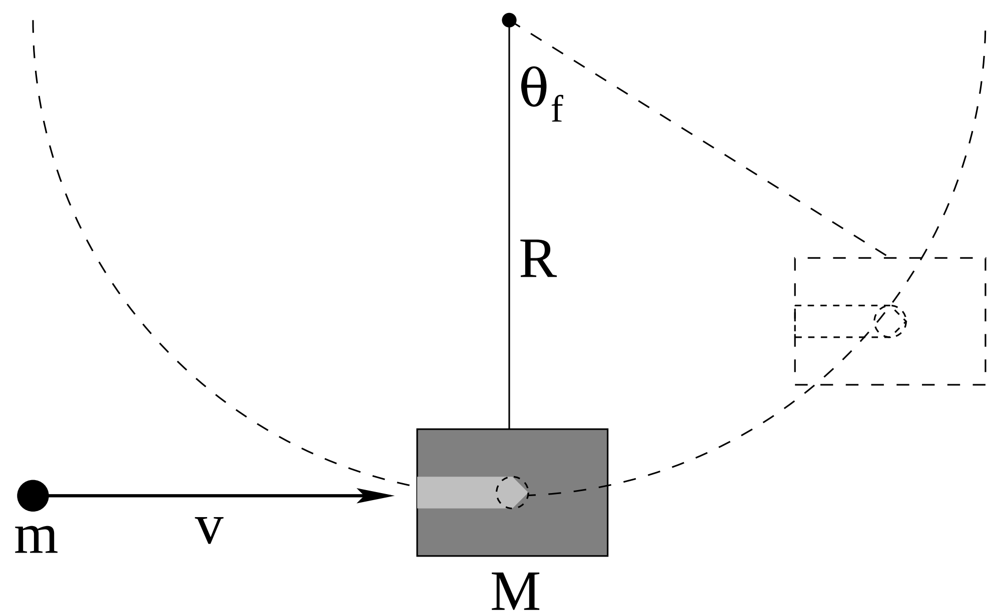


Figure 4.18: The "ballistic pendulum", where a bullet strikes and sticks to/in a block, which then swings up to a maximum angle θ_f before stopping and swinging back down.

The classic ballistic pendulum question gives you the mass of the block M, the mass of the bullet m, the length of a string or rod suspending the "target" block from a free pivot, and the initial velocity of the bullet v_0. It then asks for the maximum angle θ_f through which the pendulum swings after the bullet hits and sticks to the block (or alternatively, the maximum height H through which it swings). Variants abound – on your homework you might be asked to find the *minimum speed* v_0 the bullet must have in order the the block whirl around in a circle on a never-slack string, or on the end of a rod. Still other variants permit the bullet to pass through the block and emerge with a different (smaller) velocity. You should be able to do them all, if you completely understand this example (and the other physics we have learned up to now, of course).

There is an actual lab that is commonly done to illustrate the physics; in this lab one typically measures the maximum horizontal displacement of the block, but it amounts to the same thing once one does the trigonometry.

The solution is simple:

- **During the collision momentum is conserved** in the impact approximation, which in this case basically implies that the block has no time to swing up appreciably "during" the actual collision.

- **After the collision mechanical energy is conserved**. Mechanical energy is *not* conserved during the collision (see solution above of straight up inelastic collision).

One can replace the second sub-problem with *any other problem* that requires a knowledge of either v_f or K_f immediately after the collision as its initial condition. Ballistic loop-the-loop problems are entirely possible, in other words!

At this point the algebra is almost anticlimactic: The collision is one-dimensional (in the x-direction). Thus (for block M and bullet m) we have momentum conservation:

$$p_{m,0} = mv_0 = p_{M+m,f} \tag{4.132}$$

Now if we were foolish we'd evaluate $v_{M+m,f}$ to use in the next step: mechanical energy conservation. Being smart, we instead do the kinetic part of mechanical energy conservation in terms of momentum:

$$\begin{aligned} E_0 = \frac{p_{B+b,f}^2}{2(M+m)} &= \frac{p_{b,0}^2}{2(M+m)} \\ &= E_f = (M+m)gH \\ &= (M+m)gR(1-\cos\theta_f) \end{aligned} \tag{4.133}$$

Thus:

$$\theta_f = \cos^{-1}(1 - \frac{(mv_0)^2}{2(M+m)^2 gR}) \tag{4.134}$$

which only has a solution if mv_0 is less than some maximum value. What does it mean if it is greater than this value (there is no inverse cosine of an argument with magnitude bigger than 1)? Will this answer "work" if $\theta > \pi/2$, for a string? For a rod? For a track?

Don't leave your common sense at the door when solving problems using algebra!

Example 4.8.3: Partially Inelastic Collision

Let's briefly consider the previous example in the case where the bullet passes *through* the block and emerges on the far side with speed $v_1 < v_0$ (both given). How is the problem going to be different?

Not at all, not really. Momentum is still conserved during the collision, mechanical energy after. The *only two differences* are that we have to evaluate the speed v_f of the block M after the collision from *this* equation:

$$p_0 = m_1 v_0 = Mv_f + mv_1 = p_m + p_1 = p_f \tag{4.135}$$

so that:

$$v_f = \frac{\Delta p}{M} = \frac{m(v_0 - v_1)}{M} \tag{4.136}$$

We can read this as "the momentum transferred to the block is the momentum lost by the bullet" because momentum is conserved. Given v_f of the block *only*, you should be able to find e.g. the kinetic energy lost in this collision or θ_f or whatever in any of the many variants involving slightly different "after"-collision subproblems.

4.9: Kinetic Energy in Lab vs Center of Mass Frame

Finally, let's consider the relationship between kinetic energy in the lab frame and the CM frame, using all of the velocity relations we developed above as needed. We start with the total kinetic energy of a collection of particles in the lab frame:

$$K_{\rm tot} = \sum_i \frac{1}{2} m_i v_i^2 = \sum_i \frac{1}{2} m_i \vec{\boldsymbol{v}}_i \cdot \vec{\boldsymbol{v}}_i \tag{4.137}$$

in the lab/rest frame.

We recall (from above) the relationship between $\vec{\boldsymbol{v}}$ in the lab frame and $\vec{\boldsymbol{v}}'$ in the center of mass frame (the velocity *in* the frame plus the velocity *of* the frame):

$$\vec{\boldsymbol{v}}_i = \vec{\boldsymbol{v}}_i' + \vec{\boldsymbol{v}}_{\rm cm} \tag{4.138}$$

Then:

$$\begin{aligned} K_{\rm tot} &= \sum_i \frac{1}{2} m_i (\vec{\boldsymbol{v}}_i' + \vec{\boldsymbol{v}}_{\rm cm}) \cdot (\vec{\boldsymbol{v}}_i' + \vec{\boldsymbol{v}}_{\rm cm}) = \sum_i \frac{1}{2} m_i (v_i'^2 + v_{\rm cm}^2 + 2\vec{\boldsymbol{v}}_{\rm cm} \cdot \vec{\boldsymbol{v}}_i') \\ &= \left(\sum_i \frac{1}{2} m_i {v_i'}^2 \right) + \frac{1}{2} \left(\sum_i m_i \right) v_{\rm cm}^2 + \vec{\boldsymbol{v}}_{\rm cm} \cdot \underbrace{\left(\sum_i m_i \vec{\boldsymbol{v}}'_i \right)}_{= \sum_i \vec{\boldsymbol{p}}'_i \ = 0} \end{aligned} \tag{4.139}$$

Note well! The total momentum *in* the center of mass frame is by construction ***zero***, as the velocity *of* the center of mass frame *in* the center of mass frame is obviously zero.

Finally, sum up and identify the two surviving pieces:

$$K_{\rm tot} = \sum_i K_i = \left(\sum_i \frac{1}{2} m_i v_i' 2 \right) + \frac{1}{2} M_{\rm tot} v_{\rm cm}^2 \tag{4.140}$$

or (identifying the term inside the parentheses as the total kinetic energy of the particles in terms of their center of mass velocities, squared):

$$K_{\rm tot} = K(\boldsymbol{in}\ \ {\rm cm}) + K(\boldsymbol{of}\ \ {\rm cm}) \tag{4.141}$$

An excellent way to remember this is that the total kinetic energy in the lab frame is the sum of the kinetic energy of all the particles ***in*** the CM frame plus the kinetic energy ***of*** the CM frame (system) itself (viewed as single "object")!

At last we can understand the mystery of the baseball – how it behaves like a particle itself and yet also accounts for all of the myriad of particles it is made up of. The Newtonian motion of the baseball as a system of particles is identical to that of a particle of the same mass experiencing the same total force. The "best" location to assign the baseball (of all of the points inside) is the center of mass of the baseball. In the frame of the CM of the baseball, the total momentum of the parts of the baseball is zero (but the baseball itself has momentum $M_{\rm tot}\vec{\boldsymbol{v}}$ relative to the ground). Finally, the kinetic energy of a baseball flying through the air is the kinetic energy of the "baseball itself" (the entire system viewed as a particle) plus the kinetic energy of all the particles that make up the baseball measured in the CM frame of the baseball itself. This is comprised of **rotational kinetic energy** (which we will shortly treat) plus all the general vibrational (atomic) kinetic energy that is what we would call **heat**.

We see that we can indeed break up big systems into smaller/simpler systems, solve the smaller problems, and reassemble the solutions into a big solution, even as we can combine many, many small problems into one bigger and simpler problem and *ignore* or *average over* the details of what goes on "inside" the little problems. Treating many bodies at the same time can be quite complex, and we've only scratched the surface here, but it should be enough to help you understand both many things in your daily life and (just as important) the rest of this book.

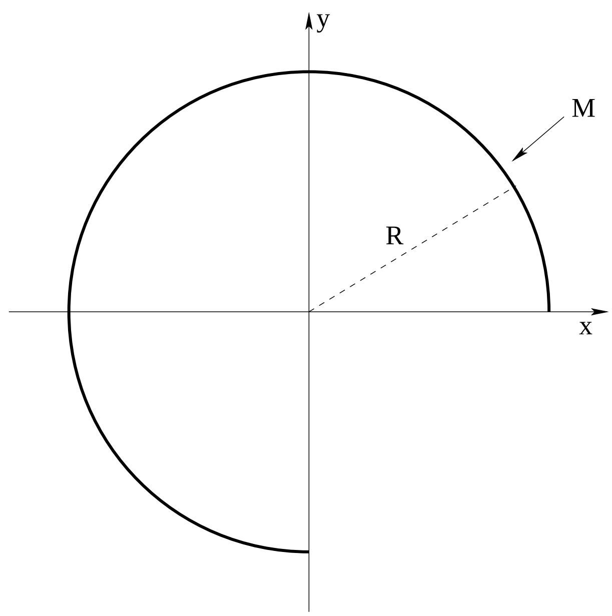
Next up (after the homework) we'll pursue this idea of motion ***in*** plus motion ***of*** a bit further in the context of ***torque*** and ***rotating systems***.

Homework for Week 4

Problem 1.

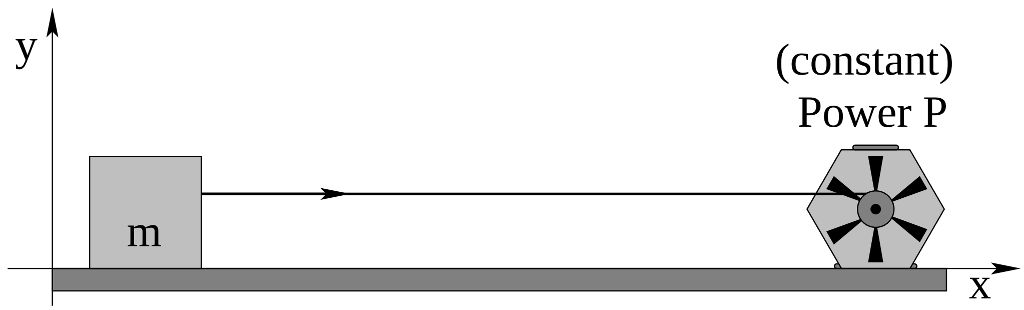
Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

Problem 2.



In the figure above, a uniformly thick piece of wire is bent into 3/4 of a circular arc as shown. Find the center of mass of the wire in the coordinate system given, using integration to find the $x_{\rm cm}$ and $y_{\rm cm}$ components separately.

Problem 3.



Suppose we have a block of mass m sitting initially at rest on a table. A massless string is attached to the block and to a motor that delivers a constant *power* P to the block as it pulls it in the x-direction.

a) Find the tension T in the string as a function of v, the speed of the block in the x-direction, ***initially*** assuming that the table is frictionless.

b) Find the acceleration of the block as a function of v.

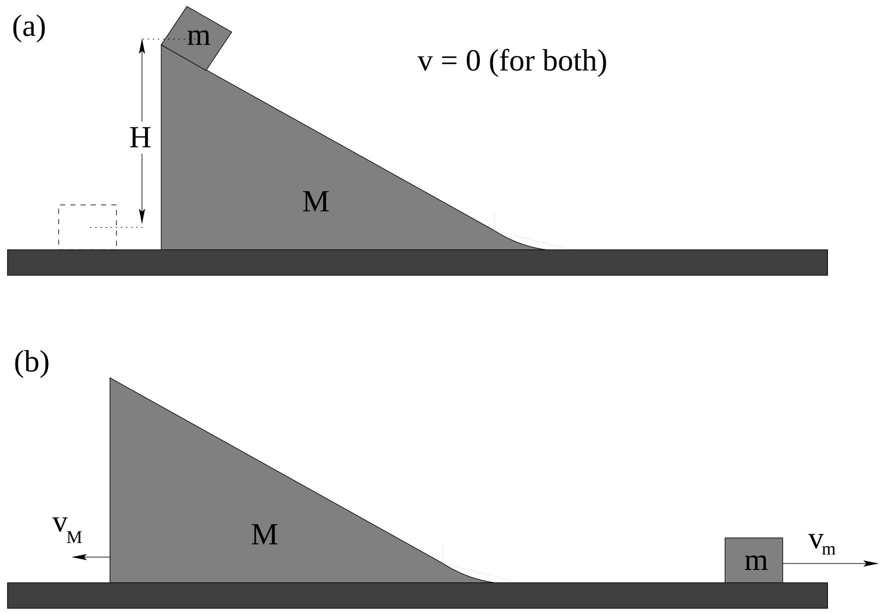
c) Solve the equation of motion to find the velocity of the block as a function of time. Show that the result is the same that you would get by evaluating:

$$\int_{t_i}^{t_f} P dt = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2$$

with $t_i = 0$, $t_f = t$, and $v_i = 0$.

d) Suppose that the table exerts a constant force of kinetic friction on the block in the opposite direction to v, with a coefficient of kinetic friction μ_k. Find the "terminal velocity" of the system after a very long time has passed. Hint: What is the ***total power*** delivered to the block by the motor and friction combined at that time?

Problem 4.



A small block with mass m is sitting on a large block of mass M that is sloped so that the small block can slide down the larger block. There is ***no friction between the two blocks, no friction between the large block and the table, and no drag force***. The *center of mass* of the small block is located a height H above where it would be if it were sitting on the table. At time $t = 0$ both blocks are ***released from rest*** (so that the *total momentum* of this system is initially *zero*, note well!)

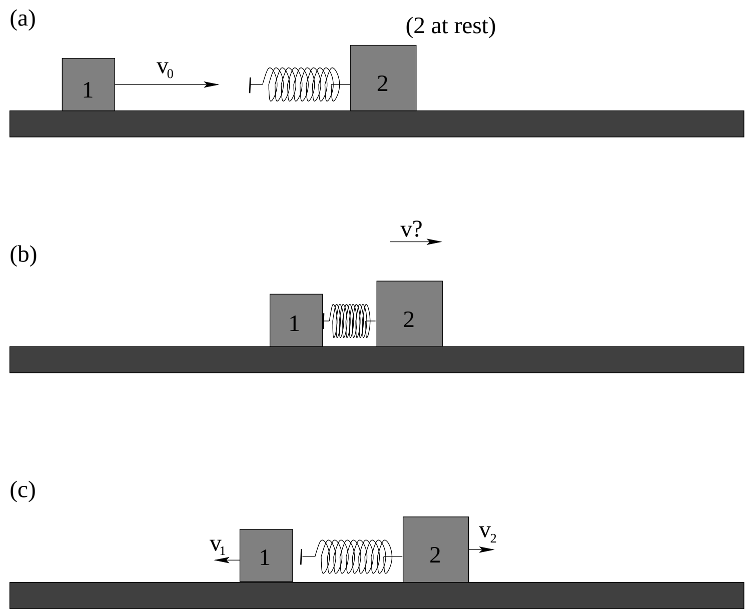
Starting hint: Are there any *net* external forces acting in this problem? In what direction do they act? What quantities do you expect to be conserved?

a) Using suitable conservation laws, find the velocities of the two blocks after the small block has slid down the larger one and they have separated.

b) To check your answer, consider the limiting case of $M \to \infty$ (where one rather expects the larger block to pretty much not move). Does your answer to part b) give you the usual result for a block of mass m sliding down from a height H on a *fixed* incline?

c) This problem doesn't *look* like a collision problem, but it easily could be half of one. Look carefully at your answer, and see if you can determine what initial velocity one should give the two blocks so that they would move *together* and precisely come to rest with the smaller block a height H above the ground. If you put the two halves together, you have solved a fully elastic collision in one dimension in the case where the center of mass velocity is zero!

Problem 5.



In (a) above, mass m_1 approaches mass m_2 at velocity v_0 to the right. Mass m_2 is initially at rest, and both sit on a frictionless table in a vacuum. An ideal massless "stiff" spring with spring constant k is attached to mass m_2.

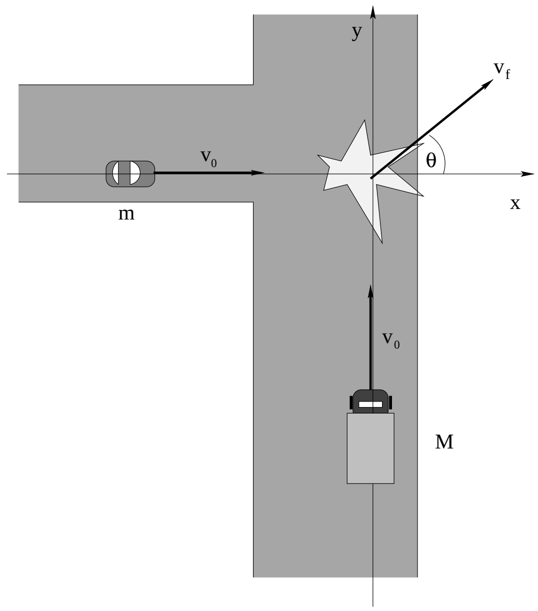
a) In panel (b) above, mass m_1 has collided with mass m_2, compressing the spring. At the particular instant shown, ***both masses are moving with the same velocity to the right***. Find and name this velocity. What physical principle do you use?

b) Find the compression Δx of the spring at this instant.

c) In panel (c) the spring has sprung back, pushing the two masses apart. Find the final velocities of the two masses.

d) In this particular problem one could in principle solve Newton's second law because the elastic collision force is *known*. In general, of course, it is not known, although for a very *stiff* spring this model is an excellent one to model elastic collisions between hard objects. Assuming that the spring is sufficiently stiff that the two masses are in contact for a very short time Δt, write a simple expression for the *impulse* imparted to m_2 and *qualitatively* sketch $F_x(t)$ and $F_{\text{avg}}(t)$ over this time interval.

Problem 6.



In the figure above, a large, heavy truck with mass M speeds through a red light to collide with a small, light compact car of mass m. Both cars fail to brake and are travelling at the speed limit (v_0) at the time of the collision, and their metal frames tangle together in the collision so that after the collision they move as one big mass[127]

a) Which exerts a larger force on the other, the car or the truck?

b) Which transfers a larger momentum to the other, the car or the truck?

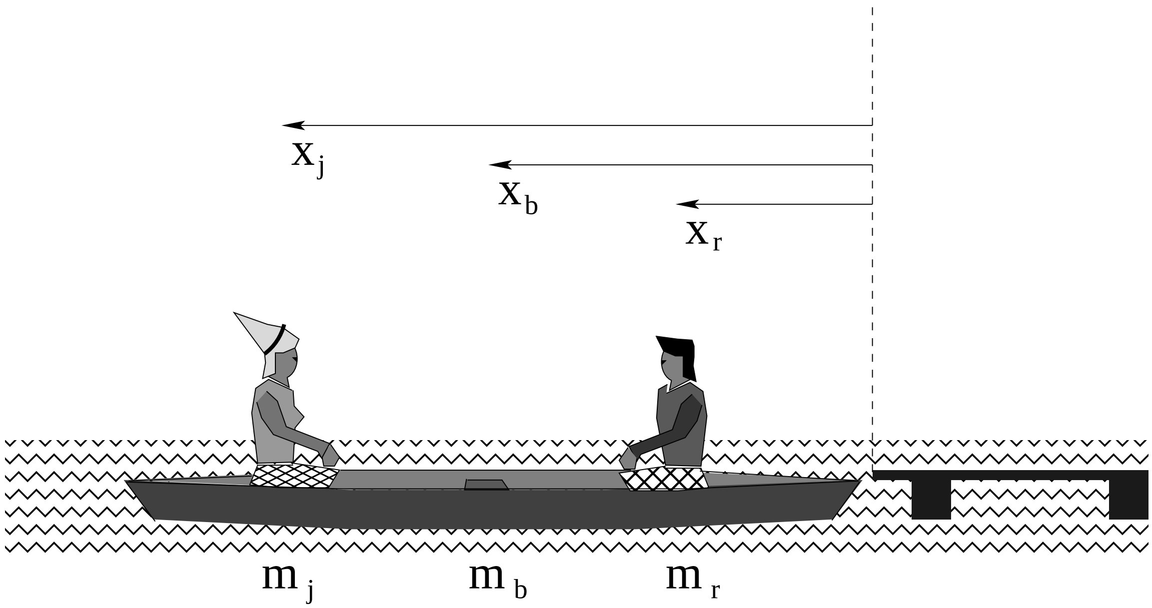
c) What is the final velocity of the wreck immediately after the collision (please give (v_f, θ))?

d) How much kinetic energy was lost in the collision?

e) If the tires blow and the wreckage has a coefficient of kinetic friction μ_k with the ground after the collision, how far D does the wreck slide before coming to rest?

[127] Both vehicles are being driven by roboticized crash test dummies. No humans were killed or injured in constructing and testing this problem! Well, except for the guy that got a paper cut on his finger while drawing out the scenario...

Problem 7.



Romeo and Juliet are sitting in a boat at rest next to a dock, looking deeply into each other's eyes. Juliet, overcome with emotion, walks from her end of the boat to sit beside Romeo and give him a chaste kiss on the cheek. The water exerts negligible friction or drag force on the boat along its length as she moves.

Assume that the masses and initial positions of the centers of mass of Romeo, Juliet and the boat are (m_r, x_r), (m_j, x_j), (m_b, x_b), (where x is measured from the dock as shown). (**Hints:** What is the velocity of the center of mass given your answer to a)? Did the center of mass move? Draw a good picture of the boat with Juliet sitting with Romeo on the same coordinate frame! Note that none of the answers below depend on whether Juliet walks at a constant speed! She might pause half way, overcome with emotion, and then continue!)

a) Given no x-directed friction or drag forces, what quantity is conserved while she is moving?

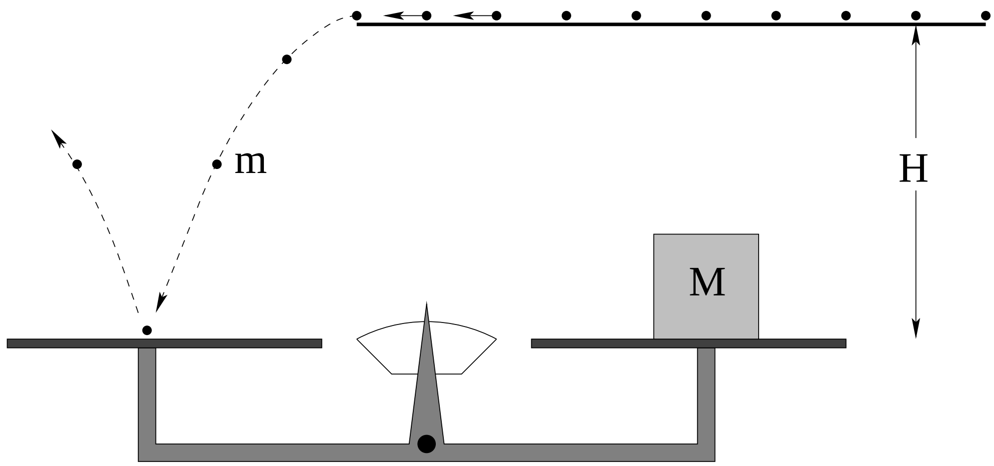
b) How far D has the boat moved away from the dock when she reaches him?

c) Does your answer to b) make sense when $m_b \gg m_j + m_r$ (the boat is the Titanic, for example) and when $m_j \gg m_b + m_r$ (the boat is an ultra-light canoe and Romeo is a tiny Romeo-doll)?

d) While she is moving to the right at the instantaneous speed v, the boat and Romeo are moving at speed v' in the opposite direction. What is the ratio v'/v?

e) What then is their *relative* speed of approach $v_{\rm rel} = v - v'$ in terms of v?

Problem 8.



In the figure above, a feeder device provides a steady stream of beads of mass m that fall a distance H and bounce elastically off of one of the hard metal pans of a beam balance scale (and then fall somewhere else into a hopper and disappear from our problem). N beads per second come out of the feeder to bounce off of the pan. Our goal is to derive an expression for M, the mass we should put on the other pan to balance the *average force* exerted by this stream of beads[128]

a) First, the easy part. The beads come off of the feeder with an initial velocity of $\vec{v} = v_{0x}\hat{x}$ in the x-direction only. Find the y-component of the velocity v_y when a single bead hits the pan after falling a height H.

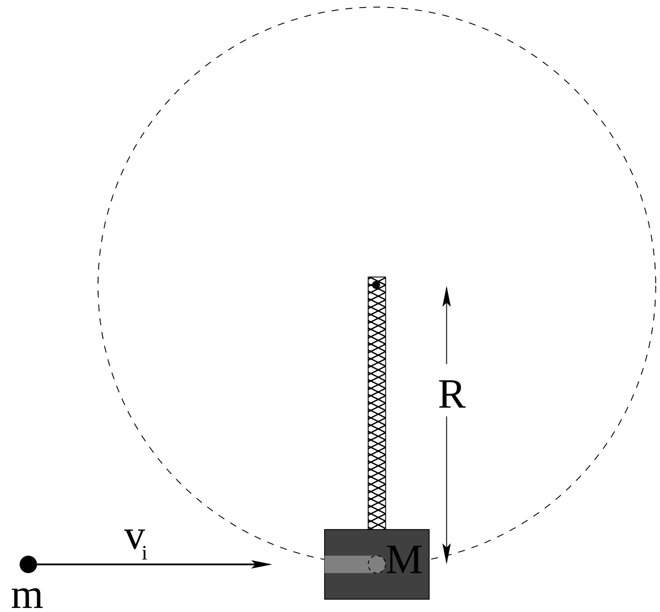
b) Since the beads bounce elastically, the x-component of their velocity is unchanged and the y-component *reverses*. Find the change of the *momentum* of this bead $\Delta\vec{p}$ during its collision.

c) Compute the *average force* being exerted on the stream of beads by the pan over a second (assuming that $N \gg 1$, so that many beads strike the pan per second).

d) Use Newton's Third Law to deduce the average force exerted by the beads on the pan, and from this determine the mass M that would produce the same force on the other pan to keep the scale in balance.

[128]This is very similar (conceptually) to the way a gas microscopically exerts a force on a surface that confines it; in a future course we will later use this idea to understand the pressure exerted by a fluid and to derive the kinetic theory of gases and the ideal gas law $PV = NkT$, which is why I assign it in particular now.

Problem 9.

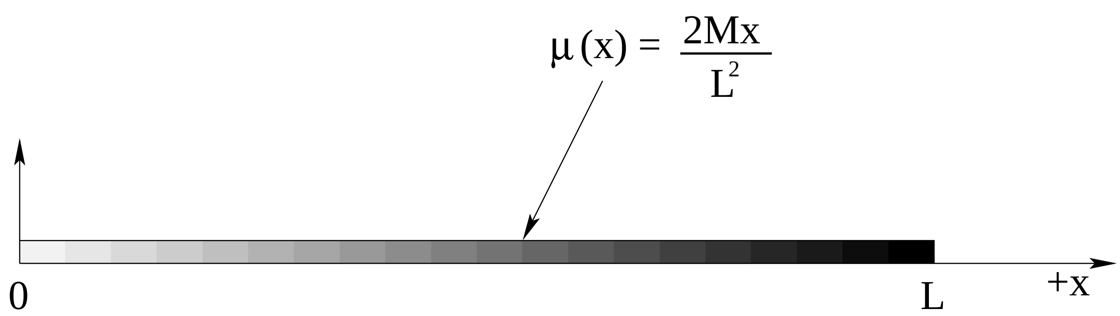


A block of mass M is attached to a rigid massless rod of length R (pivot to center-of-mass of the block/bullet distance at collision) and is suspended from a frictionless pivot. A bullet of mass m traveling at velocity v_i strikes it as shown and is quickly stopped by friction in the hole so that the two masses ***move together as one*** thereafter. Find:

a) The energy lost in the collision when the bullet is incident at this speed.

b) The minimum speed $v_{i,\mathrm{min}}$ that the bullet must have in order to swing through a complete circle after the collision. Note well that the pendulum is attached to a ***rod***!

Problem 10.



In the figure above a rod of total mass M and length L is portrayed that has been machined so that it has a mass per unit length that increases *linearly* along the length of the rod:

$$\mu(x) = \frac{2M}{L^2}x$$

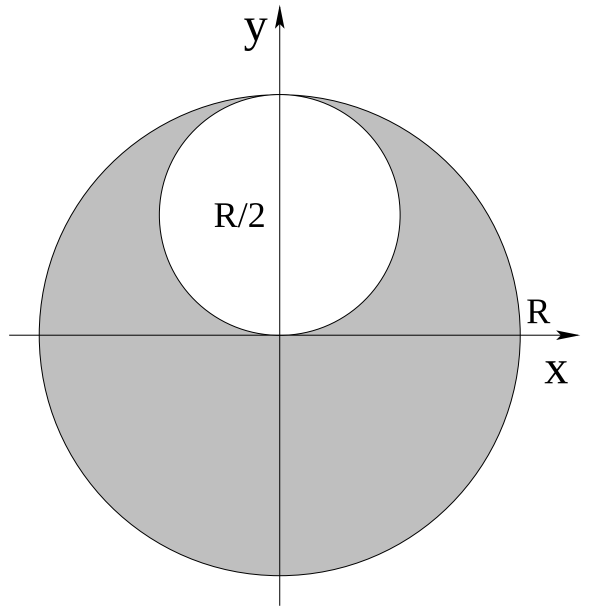
This might be viewed as a very crude model for the way mass is distributed in something like a baseball bat or tennis racket, with most of the mass near one end of a long object and very little near the other (and a continuum in between).

Treat the rod as if it is really one dimensional (we know that the center of mass will be in the center of the rod as far as y or z are concerned, but the rod is so thin that we can imagine that $y \approx z \approx 0$) and:

a) Verify that the total mass of the rod is indeed M for this mass distribution (trivial).

b) Find $x_{\rm cm}$, the x-coordinate of the center of mass of the rod.

Problem 11.

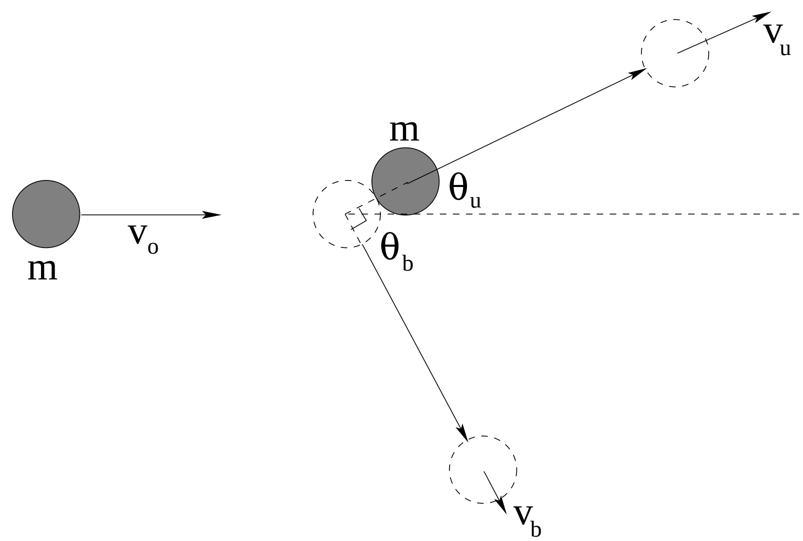


A flat disk of radius R has a circular disk of radius $R/2$ cut out of it centered at $y = R/2$ on the y-axis. The mass of the *disk that was cut out to make the hole* is M_h. Find:

a) The mass M_d of the remaining disk-with-a-hole in terms of M_h.

b) The position of the center of mass of the disk-with-a-hole. (Note that the textbook example "Square missing a square corner" is a good hint as to how to proceed here.)

Advanced Problem 12.



In the figure above, two identical billiard balls of mass m are sitting in a zero gravity vacuum (so that we can neglect drag forces and gravity). The one on the left is given a push in the x-direction so that it elastically strikes the one on the right (which is at rest) off center at speed v_0. The top ball recoils along the direction shown at a speed v_u and angle θ_u relative to the direction of incidence of the bottom ball, which is deflected so that it comes out of the collision at speed v_b at angle θ_b relative to this direction. Both balls move only *in the plane of the page* as shown, so the collision is effectively 2D.

a) Use conservation of momentum to show that in this special case that the two masses are equal:
$$\vec{v}_0 = \vec{v}_u + \vec{v}_b$$
and draw this out as a triangle.

b) Use the fact that the collision was *elastic* to show that
$$v_0^2 = v_u^2 + v_b^2$$
(where these speeds are the lengths of the vectors in the triangle you just drew).

c) Identify this equation and triangle with the *pythagorean theorem* proving that in this case $\vec{v}_t \perp \vec{v}_b$ (so that
$$\theta_u = \theta_b + \pi/2$$

Using these results, one can actually solve for $\vec{v}_u$ and $\vec{v}_b$ given only v_0 and either of θ_u or θ_b. Reasoning very similar to this is used to analyze the results of e.g. nuclear scattering experiments at various laboratories around the world (including Duke)!

Week 5: Torque and Rotation in One Dimension

1.11: Summary

- **Rotations in One Dimension** are rotations of a solid object about a *single* axis. Since we are free to choose any arbitrary coordinate system we wish in a problem, we can without loss of generality select a coordinate system where the z-axis represents the (positive or negative) direction or rotation, so that the rotating object rotates "in" the xy plane. Rotations of a rigid body in the xy plane can then be described by a *single angle* θ, measured by convention in the counterclockwise direction from the positive x-axis.

- **Time-dependent Rotations** can thus be described by:

  a) The ***angular position*** as a function of time, $\theta(t)$.

  b) The ***angular velocity*** as a function of time,

  $$\Omega(t) = \frac{d\theta}{dt}$$

  c) The ***angular acceleration*** as a function of time,

  $$\alpha(t) = \frac{d\Omega}{dt} = \frac{d^2\theta}{dt^2}$$

  Hopefully the analogy between these "one dimensional" angular coordinates and their one dimensional linear motion counterparts is obvious.

- Forces applied to a rigid object perpendicular to a line drawn from an ***axis of rotation*** exert a ***torque*** on the object. The torque is given by:

  $$\tau = rF\sin(\phi) = rF_{\perp} = r_{\perp}F$$

- The torque (as we shall see) is a *vector* quantity and by convention its direction is *perpendicular* to the plane containing $\vec{r}$ and $\vec{F}$ in the direction given by the ***right hand rule***. Although we won't really work with this until next week, the "proper" definition of the torque is:

  $$\vec{\tau} = \vec{r} \times \vec{F}$$

- ***Newton's Second Law for Rotation*** in one dimension is:

$$\tau = I\alpha$$

  where I is the ***moment of inertia*** of the rigid body being rotated by the torqe about a given/specified *axis of rotation*. The direction of this (one dimensional) rotation is the right-handed direction of the axis – the direction your right handed thumb points if you grasp the axis with your fingers curling around the axis in the direction of the rotation or torque.

- The ***moment of inertia of a point particle*** of mass m located a (fixed) distance r from some axis of rotation is:

$$I = mr^2$$

- The moment of inertia of a rigid collection of point particles is:

$$I = \sum_i m_i r_i^2$$

- the moment of inertia of a continuous solid rigid object is:

$$I = \int r^2 dm$$

- The rotational kinetic energy of a rigid body (total kinetic energy of all of the chunks of mass that make it up) is:

$$K_{\rm rot} = \frac{1}{2} I\Omega^2$$

- The work done by a torque as it rotates a rigid body through some angle $d\theta$ is:

$$dW = \tau \ d\theta$$

  Hence the work-kinetic energy theorem becomes:

$$W = \int \tau \ d\theta = \Delta K_{\rm rot}$$

- Consequently rotational work, rotational potential energy, and rotational kinetic energy call all be simply added in the appropriate places to our theory of work and energy. The total mechanical energy includes *both* the total translational kinetic energy of the rigid body treated as if it is a total mass located at its center of mass *plus* the kinetic energy of rotation *around* its center of mass:

$$K_{\rm tot} = K_{\rm cm} + K_{\rm rot}$$

  This is a special case of the last theorem we proved last week.

- If we know the moment of inertia $I_{\rm cm}$ of a rigid body about a given axis through its center of mass, the ***Parallel Axis Theorem*** permits us to find the moment of inertia of a rigid body of mass m around a *new* axis parallel to this axis and displaced from it by a distance $r_{\rm cm}$:

$$I_{\rm new} = I_{\rm cm} + mr_{\rm cm}^2$$

- For a distribution of mass with planar symmetry (mirror symmetry about the plane of rotation or distribution only in the plane of rotation), if we let z point in the direction of an axis of rotation perpendicular to this plane and x and y be perpendicular axes in the plane of rotation, then the ***Perpendicular Axis Theorem*** states that:

$$I_z = I_x + I_y$$

5.1: Rotational Coordinates in One Dimension

In the last week/chapter, you learned how a collection of particles can behave like a "particle" of the same total mass located at the center of mass as far as Newton's Second Law is concerned. We also saw at least four examples of how problems involving systems of particles can be decomposed into two separate problems – one the motion ***of*** the center of mass, which generally obeys Newtonian dynamics as if the whole system is "a particle", and the other the motion ***in*** the center of mass system[129].

This decomposition is useful (as we saw) even if the system has many particles in it and is fluid or non-interacting, but it is *very* useful in helping us to describe the ***motion of rigid bodies***. This is because the most general motion of a rigid object is the ***translation*** of (the center of mass of) the object according to the *total* force acting on it and Newton's Second Law (as demonstrated last week), plus the *rotation* of that body about its center of mass as unbalanced forces exert a ***torque*** on the object.

The first part we are very very familiar with at this point and we'll take it for granted that you can solve for the motion of the center of mass of a rigid object given any reasonable net force. The second we are not familiar with at all, and we will now take the next two weeks to study it in detail and understand it, as rotation is ***just as important and common*** as translation when it comes to understanding the motion of nearly everything we see on a daily basis. Doors rotate about hinges, tires rotate about axles, electrons and protons "just rotate" because of their intrinsic spin, our fingers and toes and head and arms and legs rotate about their joints, our whole bodies rotate about their center of mass when we get up in the morning, when we do a twirl on ice skates, when we summersault on a trampoline, the entire Earth rotates around its axis while revolving around the sun which rotates on *its* axis while revolving around the Galactic center which... just goes to show that rotation really is ubiquitous, and pretending that it isn't important or worthy of understanding is not an option, even for future physicians or non-rocket-scientist bio majors.

It will *take* two weeks (and maybe even longer, for some of you) because rotation is a wee bit *complicated*. For many of you, it will be the most difficult single topic we cover this semester, if only because rotation is best described by means of the Evil Cross Product[130] . Just as we

[129] In particular, we solved elastic collisions in the center of mass frame (where they were easy) while the center of mass of the colliding system obeyed (trivial) Newtonian dynamics, we looked at the exploding rocket where the center of mass followed the parabolic/Newtonian trajectory, we saw that inelastic collisions turn all of the kinetic energy *in* the center of mass frame into heat, and we proved that in general the kinetic energy of a system in the lab is the sum of the kinetic energy *of* the system (treated as a particle moving at speed $v_{\rm cm}$) plus the kinetic energy of all of the particles *in* the center of mass frame – this latter being the energy lost in a completely inelastic collision or conserved in an elastic one!

[130] Wikipedia: http://www.wikipedia.org/wiki/Cross Product. Something that is covered both in this Wikipedia arti-

started our study of coordinate motion with motion in only one dimension, so we will start our study or rotation with "one dimensional rotation" of a rigid body, that is, the rotation of a rigid object through an angle θ about a single fixed axis[131].

Eventually we want to be able to treat arbitrary rigid objects, ones that have their mass symmetrically but non-uniformly distributed (e.g. basketballs or ninja stars) or non-uniformly and not particularly symmetrically distributed (e.g. the human body, automobiles, blobs of putty of arbitrary shape). But at the moment even the rotation of a basketball on the tip of a player's finger seems like too much for us to handle

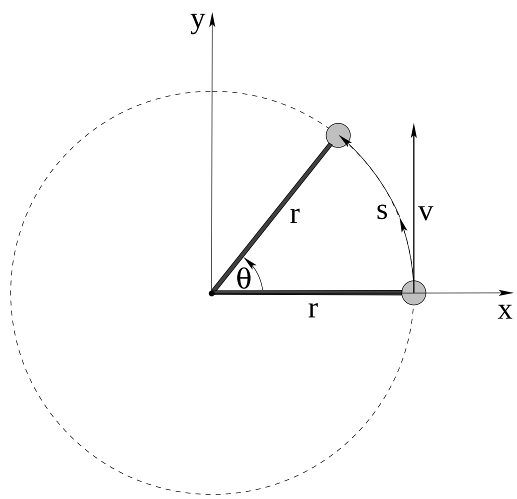


Figure 5.1: A small ball of mass m rotates about a frictionless pivot, moving in a circle of radius r.

We therefore start with the simplest possible example – a "rigid" system with all of its mass concentrated in a single point that rotates around some fixed axis. Consider a small "pointlike" ball of mass m on a *rigid* massless unstretchable rod, portrayed in figure 5.1. The rod itself is pivoted on a frictionless axle in the center so that the mass is constrained to move *only* on the dashed circle in the plane of the picture. The mass therefore maintains a constant *distance* from the pivot – r is a constant – but the *angle* θ can vary in time as external forces act on the system.

The very first things we need to do are to bring to mind the set of *rotational coordinates* that we have already introduced for doing kinematics of a rotating object. Since r is fixed, the position of the particle is uniquely determined by the positive angle $\theta(t)$, measured by convention as a counterclockwise rotation about the z-axis from the $+x$-axis as drawn in figure **??**. We call θ the ***angular position*** of the particle.

We can easily relate r and θ to the real position of the particle. The distance the particle must move in the counterclockwise direction from the standard reference position at $(x = r, y = 0)$ around the circular arc to an arbitrary position on the circle is $s = r\theta$. s (the arc length) is

cle *and* in the online Math Review supplement, so now is a really, really great time to pause in reading this chapter and skip off to refresh your memory of it. It *is* a memory, we hope, isn't it? If not, then by all means skip off to *learn* it...

[131]The "direction" of a rotation is considered to be along the axis of its rotation in a right handed sense described later below. So a "one dimensional rotation" is the rotation of any object about a single axis – it does not imply that the *object being rotated* is in any sense one dimensional.

a *one dimensional coordinate* that describes its motion on the arc of the circle itself, and if we know r and s (the latter measured from the $+x$-axis) we know exactly where the particle is in the x-y plane.

We recall that the *tangential velocity* of the particle on this circle is then

$$v_t = \frac{ds}{dt} = \frac{d(r\theta)}{dt} = r\frac{d\theta}{dt} = r\Omega \tag{5.1}$$

where we remind you of the ***angular velocity*** $\Omega = \frac{d\theta}{dt}$. Note that for a rigid body $v_r = \frac{dr}{dt} = 0$, that is, the particle is *constrained* by the rigid rod or solidity of the body to move in circles of constant radius r about the center of rotation or pivot so its speed moving towards or away from the circle is zero.

Similarly, we can differentiate one more time to find the *tangential acceleration*:

$$a_t = \frac{dv_t}{dt} = r\frac{d\Omega}{dt} = r\frac{d^2\theta}{dt^2} = r\alpha \tag{5.2}$$

where $\alpha = \frac{d\Omega}{dt} = \frac{d^2\theta}{dt^2}$ is the ***angular accleration*** of the particle.

Although the *magnitude* of $v_r = 0$, we note well that the *direction* of $\vec{v}_t$ is constantly changing and we know that $a_r = -v^2/r = -r\Omega^2$ which we derived in the first couple of weeks and by now have used repeatedly to solve many problems.

All of this can reasonably be put in a small table that lets us compare and contrast the one dimensional arc coordinates with the associated angular coordinates:

Angular	Arc Length
θ	$s = r\theta$
$\Omega = \frac{d\theta}{dt}$	$v_t = \frac{ds}{dt} = r\Omega$
$\alpha = \frac{d\Omega}{dt}$	$a_t = \frac{dv_t}{dt} = r\alpha$

Table 3: Coordinates used for angular/rotational kinemetics in one dimension. Note that θ is the rotation angle *around* a given fixed axis, in our picture above the z-axis, and that θ must be given in (dimensionless) ***radians*** for these relations to be true. Remember $C = 2\pi r$ is the formula for the circumference of a circle and is a *special case* of the general relation $s = r\theta$, but only when $\theta = 2\pi$ *radians*.

5.2: Newton's Second Law for 1D Rotations

With these coordinates in hand, we can now consider the ***angular version*** of Newton's Second Law for a force $\vec{F}$ applied to this particle as portrayed in figure 5.2. This is an example of a "rigid" body rotation, but because we aren't yet ready to tackle extended objects all of the *mass* is concentrated in the ball at radius r. We'll handle true, extended rigid objects shortly, once we understand a few basic things well.

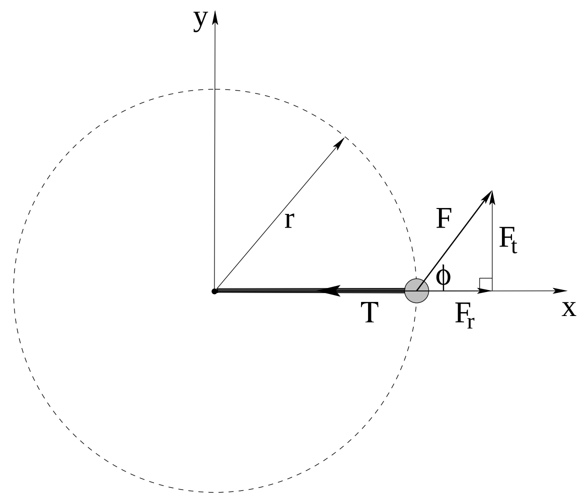


Figure 5.2: A force $\vec{F}$ is applied at some angle ϕ (relative to $\vec{r}$) to the ball on the pivoted massless rod.

Since the rod is rigid, and pivoted by an unmovable frictionless axle of some sort in the center, the tension in the rod ***opposes any motion along r***. If the particle is moving around the circle at some speed v_t (not shown), we expect that:

$$F_r - T = F\cos(\phi) - T = -ma_r = -m\frac{v_t^2}{r} = -mr\Omega^2 \tag{5.3}$$

(where r is an *outward* directed radius, note that the acceleration is *in* towards the center) as usual.

The rotational motion is what we are really interested in. Newton's Law tangent to the circle is just:

$$F_t = F\sin(\phi) = ma_t = mr\alpha \tag{5.4}$$

For reasons that will become clear in a bit, we will find it very useful to multiply this whole equation by r and redefine rF_t to be a new quantity called the ***torque***, given the symbol $\boldsymbol{\tau}$. We will also collect the factors of r and multiply them by the m to make a new quantity called the ***moment of intertia*** and give it the symbol $\boldsymbol{I}$:

$$\tau = rF_t = rF\sin(\phi) = mr^2\alpha = I\alpha \tag{5.5}$$

In particular, this equation contains the ***moment of inertia of a point mass*** m ***moving in a circle of radius*** r:

$$I_{\text{point mass}} = mr^2 \tag{5.6}$$

This looks like, and of course is, ***Newton's Second Law for a rigid rotating system*** in one dimension, where force is replaced by torque, mass is replaced by moment of inertia, and linear acceleration is replaced by angular acceleration.

Although to us *so far* this looks just like a trivial algebraic rewrite of something we could have worked with just as easily as the real thing in the s coordinates, it is actually far more general and powerful. To completely understand this, we need to understand two things. One of them is how applying the force $\vec{F}$ to (for example) the rod at *different radii* r_F changes the angular acceleration α. The other is how a force $\vec{F}$ applied at some radius r_F to the massless rod *internally redistributes* to muliple masses attached to the rod at different radii so that all the masses experience the *same* angular acceleration. These are the subjects of the next two sections.

5.2.1: The r-dependence of Torque

Let's see how the angular acceleration of this mass will scale with the point of application of the force along the rod, and in the process justify our "inspired decision" to multiply F_t by r in our definition of the torque in the previous section. To accomplish this we need a *new* figure, one where the massless rigid rod extends out past/through the mass m so it can act as a *lever arm* on the mass no matter where we choose to apply the force $\vec{F}$.

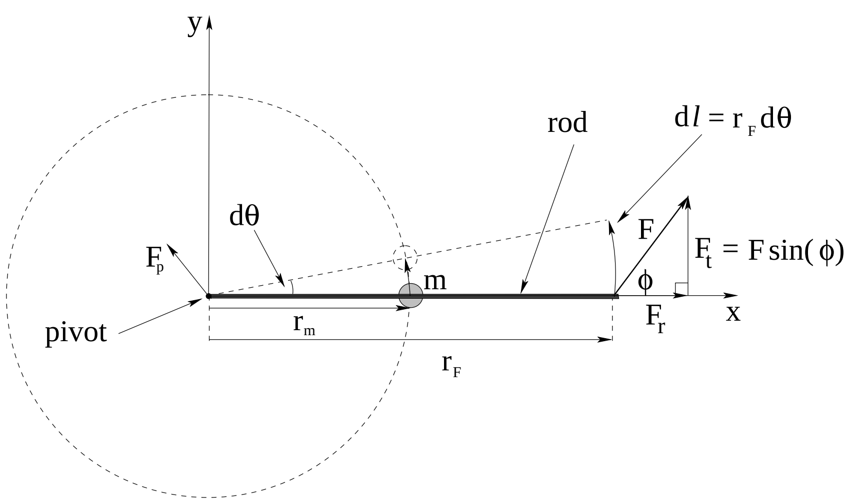


Figure 5.3: The force $\vec{F}$ is applied to the pivoted rod at an angle ϕ at the point $\vec{r}_F$ with the mass m attached to the rod at radius r_m.

This is displayed in figure 5.3. A massless rod as long or longer than both r_m and r_F is pivoted at one end so it can swing freely (no friction). The mass m is attached to the rod at the position r_m. A force $\vec{F}$ is applied to the rod at the position $\vec{r}_F$ (on the rod) and at an angle ϕ with respect to the direction of $\vec{r}_F$.

Turning this into a suitable angular equation of motion is a bit of a puzzle. The force $\vec{F}$ is not applied directly to the *mass* – it is applied to the *massless rigid rod* which in turn *transmits* some of the force to the mass. However, the external force $\vec{F}$ is not the only force acting on the rod!

In the previous example the pivot only exerted a *radial* force $\vec{F}_p = -T$, and exerted no tangential force on m at all. We could even compute T (and hence $\vec{F}_p$) if we knew θ, v_t and $\vec{F}$ from rotational kinematics and some vector geometry. In this case, however, if $\vec{F}$ exerts a force on the rod that can be transmitted to and act tangentially upon the mass m, it rather seems that the unknown pivot force $\vec{F}_p$ can as well, but ***we don't know $\vec{F}_p$!***

Alas, without knowing *all* of the forces that act tangentially on m, we cannot use Newton's Second Law *directly*. This motivates us to consider using *work and energy* to obtain a dynamical principle (basically working the derivation of the WKE theorem backwards) because ***the***

pivot does not move and therefore ***the force $\vec{F}_p$ does no work!*** Consequently, the fact that we do not *yet* know $\vec{F}_p$ *will not matter!*

So to work. Let us suppose that the force $\vec{F}$ is applied to the rod for a time dt, and that during that time the rod rotates through an angle $d\theta$ as shown. In this case we can easily find the *work* done by the force $\vec{F}$. The point on the rod where the force $\vec{F}$ is applied moves a distance $d\ell = r_F\ d\theta$. The work is done only by the tangential *component* of the force moved through this distance F_t so that:

$$dW = \vec{F} \cdot d\vec{\ell} = F_t r_F\ d\theta \tag{5.7}$$

The WKE theorem tells us that this work must equal the change in the kinetic energy over that time:

$$F_t r_F\ d\theta = dK = \frac{d\left(\frac{1}{2}mv_t^2\right)}{dt}\ dt = mv_t\ dv_t \tag{5.8}$$

We make a few useful substitutions from table 3 above:

$$F_t r_F\ d\theta = m\left(r_m \frac{d\theta}{dt}\right)(r_m \alpha\ dt) = mr_m^2 \alpha d\theta \tag{5.9}$$

and cancel $d\theta$ (and reorder a bit) to get:

$$\tau = r_F F_t = r_F F \sin(\phi) = mr_m^2 \alpha = I\alpha \tag{5.10}$$

This formally proves that my "guess" of $\tau = I\alpha$ as being the correct form of Newton's Second Law for this simple rotating rigid body holds up pretty well no matter where we apply the force(s) that make up the torque, as long as we define the torque:

$$\tau = r_F F_t = r_F F \sin(\phi) \tag{5.11}$$

It is left as an exercise for the student to draw a picture like the one above but involving *many independent and arbitrary* forces, $\vec{F}_1$ acting at $\vec{r}_1$, $\vec{F}_2$ acting at $\vec{r}_2$, ..., you get:

$$\tau_{\text{tot}} = \sum_i r_i F_i \sin(\phi_i) = mr_m^2 \alpha = I\alpha \tag{5.12}$$

for a single point-like mass on the rod at position $\vec{r}_m$. Note well that each ϕ_i is the *angle* between $\vec{r}_i$ and $\vec{F}_i$, and you should make the (massless, after all) rod long enough for all of the forces to be able to act on it and also pass through m.

In a bit we will pay attention to the fact that $rF\sin(\phi)$ is the magnitude of the ***cross product***[132] of $\vec{r}_F$ and $\vec{F}$, and that if we assign the *direction* of the rotation to be parallel to the z-axis of a right handed coordinate system when ϕ is drawn in the sense shown, we can even make this a *vector* relation. For the moment, though, we will stick with our simple 1D "scalar" formulation and ask a *different* question: what if we have a *more complicated* object than a single mass on a pivoted rigid rod that is being driven by a torque (or sum of torques).

The answer is: We have to sum up the object's total ***moment of inertia*** around the pivot axis. Let's prove this.

[132]Wikipedia: http://www.wikipedia.org/wiki/Cross Product. Making this a gangbusters good time to go review – or learn – cross products, at least well enough to be able evaluate their *magnitude* and *direction* (using the right hand rule).

5.2.2: Summing the Moment of Inertia

Suppose we have a mass m_1 attached to our massless rod pivoted at the origin at the position r_1, and a second mass m_2 attached at position r_2. We will then apply the force $\vec{F}$ at an angle ϕ to the (extended) rod at position r as shown in figure **??**, and duplicate our reasoning from the last chapter (because we *still* do not known the unknown force exerted by the pivot, but as long as we consider work we don't have to.

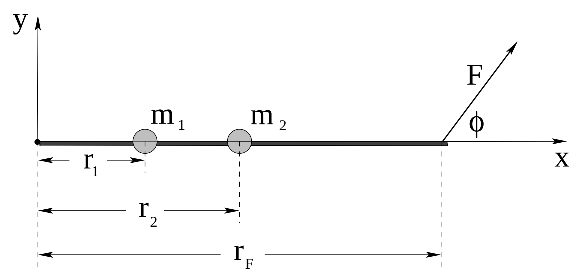


Figure 5.4: A single torque $\tau = r_F F \sin(\phi)$ is applied to a rod with *two* masses, m_1 at r_1 and m_2 at r_2.

The WKE theorem for this picture is now (note that v_1 and v_2 are both necessarily tangential):

$$\begin{aligned}
dW = r_F F \sin(\phi)\, d\theta = \tau\, d\theta &= dK = m_1 v_1\, dv_1 + m_2 v_2 dv_2 \qquad \text{so as usual} \\
\tau\, d\theta &= m_1 \left(r_1 \frac{d\theta}{dt} \right) (r_1 \alpha\, dt) + m_2 \left(r_2 \frac{d\theta}{dt} \right) (r_2 \alpha\, dt) \\
\tau &= m_1 r_1^2 \alpha + m_2 r_2^2 \alpha \\
\tau &= \left(m_1 r_1^2 + m_2 r_2^2 \right) \alpha = I\alpha
\end{aligned} \tag{5.13}$$

where we have now *defined*

$$I = m_1 r_1^2 + m_2 r_2^2 \tag{5.14}$$

That is, the total moment of inertia of the *two* point masses is just the sum of their individual moments of inertia. From the derivation it should be clear that if we added 3,4,...,N point masses along the massless rod the total moment of inertia would just be the sum of their individual moments of inertia.

Indeed, as we add more forces acting at different points and directions (in the plane) on the rod and add more masses at different points on the rod, *everything* we did above clearly *scales up linearly* – we simply have to *sum* the total torque on the right hand side and *sum* the total moment of inertia on the left hand side. We therefore conclude that Newton's Second Law for a system constrained to rotate in (one dimension in a) a plane about a fixed pivot is just:

$$\tau_{\text{tot}} = \sum_i r_i F_i \sin(\phi_i) = \sum_j m_j r_j^2 \alpha = I_{\text{tot}} \alpha \tag{5.15}$$

So much for discrete forces and discrete masses. However, most rigid bodies that we experience every day are, on a coarse-grained macroscopic scale, made up of a *continuous* distribution of mass, and instead of a mythical idealized "massless rigid rod" all of this mass is glued together by means of *internal forces*.

It is pretty clear that our expression $\tau = I\alpha$ will generalize to this case where we will (probably) replace:

$$I = \sum_j m_j r_j^2 \rightarrow \int r^2 dm \tag{5.16}$$

but we will need to do just a teensy bit of work to show that this is true and extract any essential conceptual insight to be found along the way.

5.3: The Moment of Inertia

We begin with a specific example to help smooth the way.

Example 5.3.1: The Moment of Inertia of a Rod Pivoted at One End

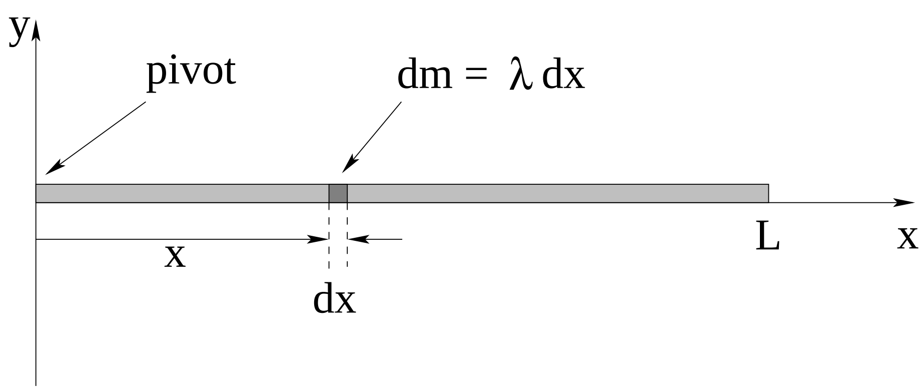


Figure 5.5: A solid rod of length L with a mass M uniformly pivoted about one end. One can think of such a rod as being the "massless rod" of the previous section with an *infinite number* of masses m_i uniformly distributed along its length, that sum to the total mass M.

In figure 5.5 above a *massive* rod pivoted about one end is drawn. We would like to determine how this particular rod will rotationally accelerate when we (for example) attach a force to it and apply a torque. We therefore must characterize this rod as having a specific mass M, a specific length L, and we need to say something about the *way* the mass is distributed, because the rod could be made of aluminum (not very dense) at one end and tungsten (very dense indeed) at the other and still "look" the same. We will assume that *this* rod is uniformly distributed, and that it is very thin and symmetrical in cross-section – shaped like a piece of wire or perhaps a wooden dowel rod.

In a process that should be familiar to you from last week and from the previous section, we know that the moment of inertia of a sum of *discrete* point masses hung on a "massless" rod (that only serves to assemble them into a rigid structure) is just:

$$I_{\text{tot}} = \sum_i m_i r_i^2 \tag{5.17}$$

the sum of the moments of inertia of the point masses.

We can clearly *approximate* the moment of inertia of the continuous rod by dividing it up into N pieces, each of length $\Delta x = L/N$ and mass $\Delta M = M/N$, and treating each small piece as a "point mass" located at $x_i = i * \Delta x$:

$$I_{\rm rod} \approx \sum_{i=1}^{N} \frac{M}{N} \left(\frac{i * L}{N} \right) = \sum_{i=1}^{N} \Delta M x_i^2 \tag{5.18}$$

As before, the limit of this sum as $N \to \infty$ is *by definition* the integral, and in this limit the sum *exactly* represents the moment of inertia of the rod.

We can easily evaluate this. To do so, we chant our ritual expression: "The mass of the chunk is the mass per unit *length* of the chunk times the length of the chunk", or $dm = \mu dx = \frac{M}{L} dx$, so:

$$I_{\rm rod} = \int_0^L x^2 dm = \frac{M}{L} \int_0^L x^2 dx = \frac{ML^2}{3} \tag{5.19}$$

5.3.1: Moment of Inertia of a General Rigid Body

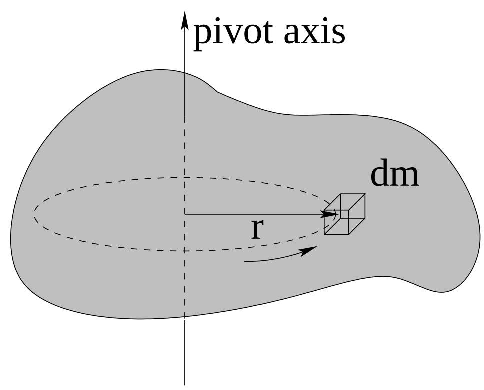


Figure 5.6: A "blob-shaped" chunk of mass, perhaps a piece of modelling clay, constrained to rotate about an axis through the blob, perhaps a straight piece of nearly massless coat-hanger wire.

This specific result can easily be generalized. If we consider a blob-shaped distribution of mass, the *differential* moment of inertia of a tiny chunk of the mass in the distribution about some fixed axis of rotation is clearly:

$$dI = r^2 dm \tag{5.20}$$

By now you should be getting the idea that summing up *all of the little chunks* that make up the object is just integrating:

$$I_{\rm blob} = \int_{\rm blob} r^2 dm \tag{5.21}$$

where it is quite one thing to write down this formal expression, quite another to be able to actually *do* the integral over all of the chunks of mass that make up an object.

It isn't too difficult to do this integral for certain *simple* distributions of mass, and we will need a certain "stock repertoire" of moments of inertia in order to solve problems. Which ones you should learn to do depends on the level of the course – math/physics majors should learn to integrate over spheres (and maybe engineers as well), but everybody else can probably get away learning to evaluate the moment of inertia of a disk. In practice, for any really *complicated* mass distribution (like the blob of clay pictured above) one would either *measure* the moment of inertia or use a computer to actually break the mass up into a very large number of discrete (but small/point-like) chunks and do the *sum*.

First let's do an example that is even simpler than the rod.

Example 5.3.2: Moment of Inertia of a Ring

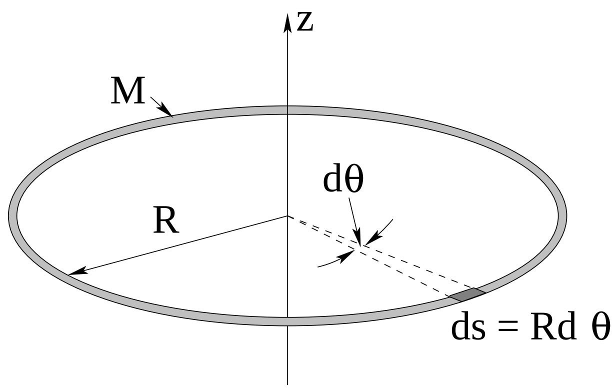


Figure 5.7: A ring of mass M and radius R in the x-y plane rotates freely about the z-axis.

We would like to find the moment of inertia of the ring of uniformly distributed mass M and radius R portrayed in figure **??** above. A differential chunk of the ring has length $ds = R\,d\theta$. It's mass is thus (say the ritual words!):

$$dm = \mu ds = \frac{M}{2\pi R} R\,d\theta = \frac{M}{2\pi}\,d\theta \tag{5.22}$$

and its moment of inertia is very simple:

$$I_{\text{ring}} = \int r^2 dm = \int_0^{2\pi} \frac{M}{2\pi} R^2\,d\theta = MR^2 \tag{5.23}$$

In fact, we could have *guessed* this. *All* of the mass M in the ring is at the same distance R from the axis of rotation, so its moment of inertia (which only depends on the mass times the distance and has no "vector" character) is just MR^2 just like a point mass at that distance.

Because it is so important, we will do the moment of inertia of a disk next. The disk will be many things to us – a massive pulley, a wheel or tire, a yo-yo, a weight on a grandfather clock (physical) pendulum. Here it is.

Example 5.3.3: Moment of Inertia of a Disk

In figure 5.8 a disk of uniformly distributed mass M and radius R is drawn. We would like to find its moment of inertia. Consider the small chunk of disk that is shaded of area dA. In plane polar coordinates (the only ones we could sanely hope to integrate over) the differential area of this chunk is just its differential height dr times the width of the arc at radius r subtended by the angle $d\theta$, $r\,d\theta$. The area is thus $dA = r\,drd\theta$.

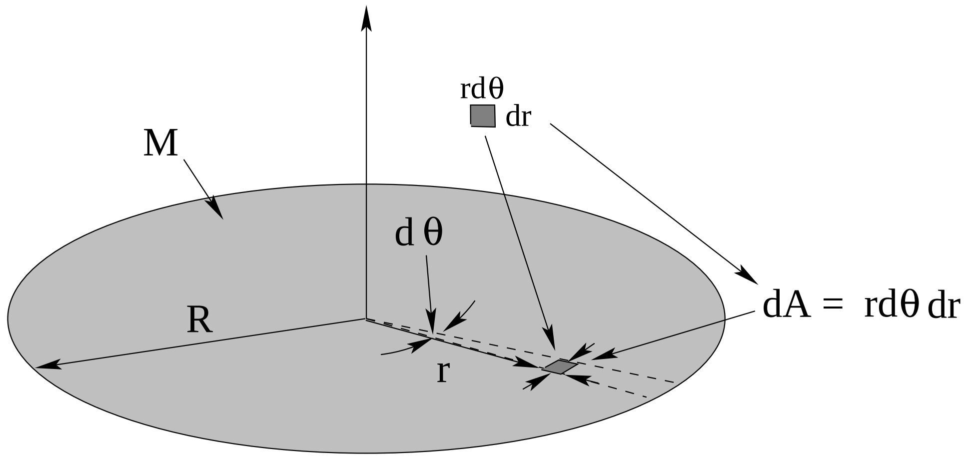


Figure 5.8: A disk of mass M and radius R is pivoted to spin freely around an axis through its center.

This little chunk was selected because the mass dm in it moves in a circle of radius r around the pivot axis. We need to find dm in units we can integrate to cover the disk. We use our litany to set:

$$dm = \sigma dA = \frac{M}{\pi R^2} r \; dr d\theta \tag{5.24}$$

and then write down:

$$dI = r^2 dm = \frac{M}{\pi R^2} r^3 \; dr d\theta \tag{5.25}$$

We integrate both sides to get (note that the integrals are independent one dimensional integrals that precisely cover the disk):

$$\begin{aligned} I_{\rm disk} &= \frac{M}{\pi R^2} \left(\int_0^R r^3 \; dr \right) \left(\int_0^{2\pi} d\theta \right) \\ &= \frac{M}{\pi R^2} \left(\frac{R^4}{4} \right) (2\pi) \\ &= \frac{1}{2} M R^2 \end{aligned} \tag{5.26}$$

This is a very important and useful result, so keep it in mind.

5.3.2: Table of Useful Moments of Inertia

Finally, here is a table of a *few* useful moments of inertia of simple uniform objects. In each case I indicate the value about an axis through the symmetric *center of mass* of the object, because we can use the ***parallel axis theorem*** and the ***perpendicular axis theorem*** to find the moments of inertia around at least some alternative axes.

5.4: Torque as a Cross Product

This section will be rather abbreviated *this* week; *next* week we will cover it in gory detail as a *vector* relation. For the moment, however, we need to make a number of observations that

Shape	$I_{\rm cm}$
Rod (about CoM)	$\frac{1}{12}ML^2$
Rod (about end)	$\frac{1}{3}ML^2$
Ring	MR^2
Disk	$\frac{1}{2}MR^2$
Sphere	$\frac{2}{5}MR^2$
Spherical Shell	$\frac{2}{3}MR^2$
Generic “Round” Mass of Mass M and radius R	βMR^2

Table 4: A few useful moments of inertia of symmetric objects around an axis of symmetry through their center of mass. You should probably know all of the moments in this table and should be able to evaluate the first three by direct integration.

will help us solve problems. First, we know that the one-dimensional torque produced by any single force acting on a rigid object a distance r from a pivot axis is just:

$$\tau = rF_\perp = rF\sin(\phi) \tag{5.27}$$

where $F_\perp$ is just the component of the force *perpendicular* to the (shortest) vector $\vec{r}$ from the pivot axis to the point of application. This is really just one *component* of the total torque, mind you, but it is the one we have learned so far and are covering this week.

First, let's make an important observation. Provided that $\vec{r}$ and $\vec{F}$ lie in a plane (so that the one dimension is the right dimension) the magnitude of the torque is the magnitude of the ***cross product of $\vec{r}$ and $\vec{F}$***:

$$\tau = |\vec{r} \times \vec{F}| = rF\sin(\phi) = rF_\perp = r_\perp F \tag{5.28}$$

I've used the fact that I can move the $\sin(\phi)$ around to write this in terms of:

$$r_\perp = r\sin(\phi) \tag{5.29}$$

which is the ***component of $\vec{r}$ perpendicular to F***, also known as the ***moment arm of the torque***. This is a very useful form of the torque in many problems. It it equally well expressible in terms of the familiar:

$$F_\perp = F\sin(\phi), \tag{5.30}$$

the component of $\vec{F}$ perpendicular to $\vec{r}$. This form, too, is often useful. In fact, both forms may be useful (to evaluate different parts of the total torque) in a single problem!

If we let the ***vector torque*** be defined by:

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{r}} \times \vec{\boldsymbol{F}} \tag{5.31}$$

marvelous things will happen. Next week we will learn about them, and will learn about how to evaluate this a variety of ways. For now let's just learn one.

> The vector torque $\vec{\boldsymbol{\tau}}$ has a *magnitude* $|\vec{\boldsymbol{r}} \times \vec{\boldsymbol{F}}| = rF\sin(\phi)$ and points in the *direction* given by the ***right hand rule***.

The right hand rule, in turn, is the following:

> The direction of the vector cross product $\vec{\boldsymbol{A}} \times \vec{\boldsymbol{B}}$ is in the direction the thumb of your ***right hand*** points when you begin with the fingers of your right hand lined up with the vector $\vec{\boldsymbol{A}}$ and then curl them naturally through the angle $\phi < \pi$ into the direction of $\vec{\boldsymbol{B}}$.

That is, if you imagine "grasping" the axis in the direction of the torque with your right hand, your fingers will curl around in the direction *from* $\vec{\boldsymbol{r}}$ *to* $\vec{\boldsymbol{F}}$ through the smaller of the two angles in between them (the one less than π).

You will get lots of practice with this rule, but be sure to practice with your *right* right hand, not your *wrong* right hand. Countless students (and physics professors and TAs!) have been embarrassed be being caught out evaluating the direction of cross products with their left hand[133]. Don't be one of them!

The direction of the torque *matters*, even in one dimension. There is no better problem to demonstrate this than the following one, determine what *direction* a spool of rope resting on a table will roll when one pulls on the rope.

Example 5.4.1: Rolling the Spool

I'm not going to *quite* finish this one for you, as there are a lot of things one can ask and it is a homework problem. But I do want you to get a good start.

The spool in figure 5.9 is wrapped many times around with string. It is sitting on a level, rough table so that for weak forces $\vec{\boldsymbol{F}}$ it will freely roll without slipping (although for a large enough $\vec{\boldsymbol{F}}$ of course it will slip or even rise up off of the table altogether).

The question is, which *direction* will it roll (or will it not roll until it slides) for each of the three directions in which the string is pulled.

The answer to this question depends on the *direction of the total torque*, and the relevant pivot is the point that *does not move* when it rolls, where the (unknown!) force of static friction acts. If we choose the pivot to be the point where the spool touches the table, then gravity, the normal force and static friction *all exert no torque!* The only source of torque is $\vec{\boldsymbol{F}}$.

So, what is the *direction* of the torque for each of the three forces drawn, and will a torque in that direction make the mass roll to the *left*, the *right*, or *slide* (or not move)?

Think about it.

[133] Let he or she who is without sin cast the first stone, I always say. As long as it is cast with the *right* hand...

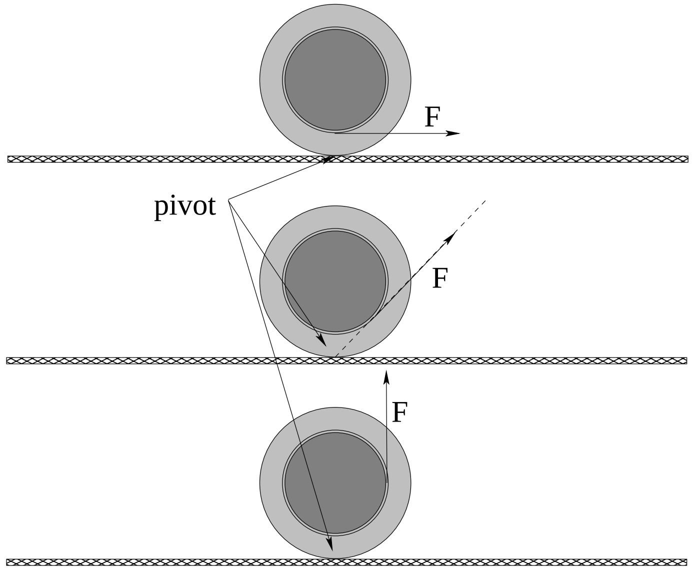


Figure 5.9: What ***direction*** does one expect the spool in each of the figures to roll (or will it roll at all)?

5.5: Torque and the Center of Gravity

We will often wish to solve problems involving (for example) rods pivoted at one end swinging down under the influence of near-Earth gravity, or a need to understand the trajectory and motion of a spinning basketball. To do this, we need the idea of the ***center of gravity*** of a solid object. Fortunately, this idea is very simple:

> The ***center of gravity*** of a solid object in an (approximately) uniform near-Earth gravitational field is located at the center of mass of the object. For the purpose of evaluating the torque and angular motion or force and coordinate motion of center of mass, we can consider that the entire force of gravity acting on the object is equivalent to the the force that would be exerted by the entire mass located as a point mass at the center of gravity.

The proof for this is very simple. We've already done the Newtonian part of it – we know that the total force of gravity acting on an object makes the center of mass move like a particle with the same mass located there. For torque, we recall that:

$$\tau = rF_\perp = r_\perp F \tag{5.32}$$

If we consider the torque acting on a small chunk of mass in near-Earth gravity, the force (down) acting on that chunk is:

$$dF_y = -g dm \tag{5.33}$$

The torque (relative to the pivot) is just:

$$d\tau = -g r_\perp dm \tag{5.34}$$

or

$$\tau = -g \int_0^L r_\perp dm = -Mg r_{\perp,\mathrm{cm}} \tag{5.35}$$

where $r_{\perp,\mathrm{cm}}$ is the component of the vector position of the center of mass of the object perpendicular to gravity. The torque due to gravity acting on the object around the selected pivot axis is the same as the torque that would be produced by the entire weight of the object pulling down at the center of mass/gravity. Q.E.D.

Note that this proof is valid for *any* shape or distribution of mass in a *uniform* gravitational field, but in a non-uniform field it would fail and the center of gravity would be *different* from the center of mass as far as computing torque is concerned. This is the realm of ***tides*** and will be discussed more in the week where we cover gravity.

Example 5.5.1: The Angular Acceleration of a Hanging Rod

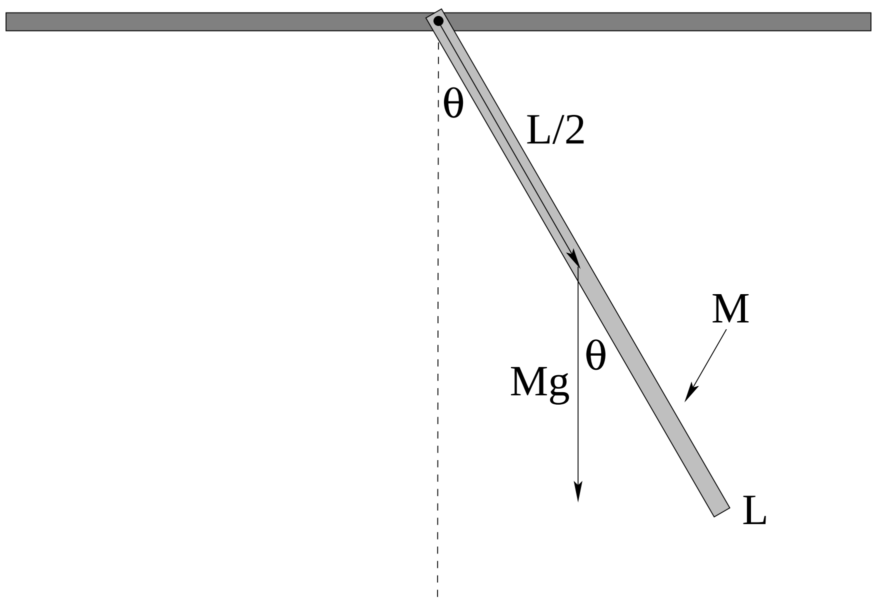


Figure 5.10: A rod of mass M and length L is suspended/pivoted from one end. It is pulled out to some initial angle θ_0 and released.

This is your first example of what we will later learn to call a ***physical pendulum***. A rod is suspended from a pivot at one end in near-Earth gravity. We wish to find the angular acceleration α as a function of θ. This is basically the equation of motion for this *rotational* system – later we will learn how to (approximately) solve it.

As shown above, for the purpose of evaluating the torque, the force due to gravity can be considered to be Mg straight down, acting at the center of mass/gravity of the rod (at $L/2$ in the middle). The torque exerted at this arbitrary angle θ (positive as drawn, note that it is swung out in the counterclockwise/right-handed direction from the dashed line) is therefore:

$$\tau = rF_t = -\frac{MgL}{2}\sin(\theta) \tag{5.36}$$

It is *negative* because it acts to make θ *smaller*; it exerts a “twist” that is clockwise when θ is counterclockwise and vice versa.

From above, we know that $I = ML^2/3$ for a rod pivoted about one end, therefore:

$$\begin{aligned}\tau &= -\frac{MgL}{2}\sin(\theta) &=& \frac{ML^2}{3}\alpha = I\alpha \qquad \text{or:} \\ \alpha &= \frac{d^2\theta}{dt^2} &=& -\frac{3g}{2L}\sin(\theta)\end{aligned} \tag{5.37}$$

independent of the mass!

5.6: Solving Newton's Second Law Problems Involving Rolling

One of the most common applications of one dimensional torque and angular momentum is solving *rolling problems*. Rolling problems include things like:

- A disk rolling down an inclined plane.
- An Atwood's Machine, but with a *massive* pulley.
- An unwinding spool of line, either falling or being pulled.

These problems are all solved by using a *combination* of Newton's Second Law for the motion of the center of mass of the rolling object (if appropriate) or other masses involved (in e.g. Atwood's Machine) *and* Newton's Second Law for 1 dimensional rotation, $\tau = I\alpha$. In general, they will also involve using the ***rolling constraint:***

> If a round object of radius r is *rolling without slipping*, the distance x it travels relative to the surface it is rolling on equals $r\theta$, where θ is the angle it rolls through.

That is, all three are equivalently the "rolling constraint" for a ball of radius r rolling on a level floor, started from a position at $x = 0$ where also $\theta = 0$:

$$x = r\theta \tag{5.38}$$
$$v = r\Omega \tag{5.39}$$
$$a = r\alpha \tag{5.40}$$
$$\tag{5.41}$$

These are all quite familiar results – they look a lot like our angular coordinate relations – but they are ***not the same thing!*** These are ***constraints***, not coordinate relations – for a ball skidding along the same floor they will be false, and for certain rolling pulley problems on your homework you'll have to figure out one appropriate for the particular radius of contact of spool-shaped or yo-yo shaped rolling objects (that may not be the radius of the object!)

It is easier to demonstrate how to proceed for specific examples than it is to expound on the theory any further. So let's do the simplest one.

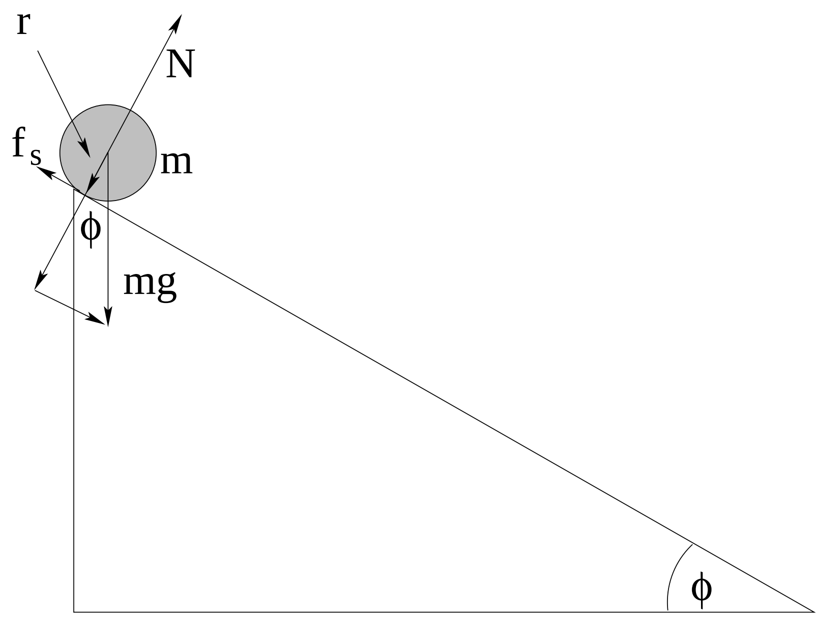


Figure 5.11: A disk of mass M and radius r sits on a plane inclined at an angle ϕ with respect to the horizontal. It ***rolls without slipping*** down the incline.

Example 5.6.1: A Disk Rolling Down an Incline

In figure 5.11 above, a disk of mass M and radius r sits on an inclined plane (at an angle ϕ) as shown. It rolls without slipping down the incline. We would like to find its acceleration a_x down the incline, because if we know that we know pretty much everything about the disk at all future times that it remains on the incline. We'd also like to know what f_s (the force exerted by static friction) when it is so accelerating, so we can check to see if our assumption of rolling without slipping is justified. If ϕ is too large, we are pretty sure intuitively that the disk will slip instead of roll, since if $\phi \geq \pi/2$ we know the disk will just fall and not roll at all.

As always, since we expect the disk to physically translate down the incline (so $\vec{a}$ and $\vec{F}_{\rm tot}$ will point that way) we choose a coordinate system with (say) the x-axis directed down the inline and y directed perpendicular to the incline.

Since the disk *rolls without slipping* we know two very important things:

1) The force f_s exerted by static friction must be less than (or marginally equal to) $\mu_s N$. If, in the end, it *isn't*, then our solution is invalid.

2) If it *does* roll, then the distance x it travels down the incline is related to the angle θ it rolls through by $x = r\theta$. This also means that $v_x = r\Omega$ and $a_x = r\alpha$.

We now proceed to write Newton's Laws ***three times***: Once for the y-direction, once for the x-direction and once for *one dimensional rotation* (the rolling). We start with the:

$$F_y = N - mg\cos(\phi) = ma_y = 0 \tag{5.42}$$

which leads us to the familar $N = mg\cos(\phi)$.

Next:

$$F_x = mg\sin(\phi) - f_s = ma_x \quad (5.43)$$
$$\tau = rf_s = I\alpha \quad (5.44)$$

Pay attention here, because we'll do the following sort of things fairly often in problems. We use $I = I_{\rm disk} = \frac{1}{2}mr^2$ and $\alpha = a_x/r$ and divide the last equation by r on both sides. This gives us:

$$mg\sin(\phi) - f_s = ma_x \quad (5.45)$$
$$f_s = \frac{1}{2}ma_x \quad (5.46)$$

If we ***add these two equations***, the unknown f_s cancels out and we get:

$$mg\sin(\phi) = \frac{3}{2}ma_x \quad (5.47)$$

or:

$$a_x = \frac{2}{3}g\sin(\phi) \quad (5.48)$$

We can then substitute this back into the equation for f_s above to get:

$$f_s = \frac{1}{2}ma_x = \frac{1}{3}mg\sin(\phi) \quad (5.49)$$

In order to roll without slipping, we know that $f_s \le \mu_s N$ or

$$\frac{1}{3}mg\sin(\phi) \le \mu_s mg\cos(\phi) \quad (5.50)$$

or

$$\mu_s \ge \frac{1}{3}\tan(\phi) \quad (5.51)$$

If μ_s is smaller than this (for any given incline angle ϕ) then the disk will *slip* as it rolls down the incline, which is a more difficult problem.

We'll solve this problem again shortly to find out how *fast* it is going at the bottom of an incline of length L using energy concepts, but in order to do this in the easiest possible way we have to consider and include *rotational* kinetic energy. First, however, we'll do another example of Newton's Law problems that mix torque and rotation with force and translation.

Example 5.6.2: Atwood's Machine with a Massive Pulley

Our solution strategy is almost identical to that of our first solution back in week 1 – choose an "around the corner" coordinate system where if we make moving m_2 down "positive", then moving m_1 up is also "positive". To this we add that a positive *rotation* of the *pulley* is *clockwise*, and that the rolling constraint is therefore $a = r\alpha$.

Now we again write Newton's Second Law once for each mass and ***once for the rotating pulley*** (as $\tau = I\alpha$):

$$m_2 g - T_2 = m_2 a \quad (5.52)$$
$$T_1 - m_1 g = m_1 a \quad (5.53)$$
$$\tau = rT_2 - rT_1 = \beta M r^2 \frac{a}{r} = I\alpha \quad (5.54)$$

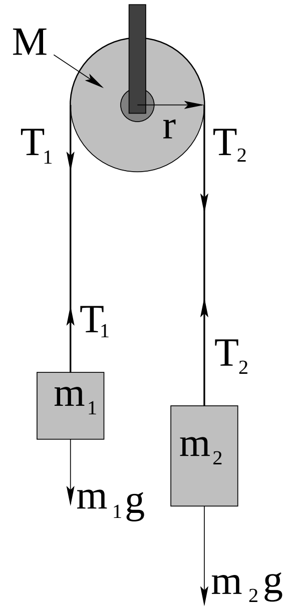


Figure 5.12: Atwood's Machine, but this time with a *massive* pulley of mass M and radius R. The massless unstretchable string connecting the two masses ***rolls without slipping*** on the pulley, exerting a torque on the pulley as the masses accelerate to match. Assume that the pulley has a moment of inertia $\beta M r^2$ for some β. Writing it this way let's us use $\beta \approx \frac{1}{2}$ to approximate the pulley with a disk, or use observations of a to *measure* β and hence tell something about the distribution of mass in the pulley!

Divide the last equation by r on both sides, then add all three equations to eliminate *both* unknown tensions T_1 and T_2. You should get:

$$(m_2 - m_1)g = (m_1 + m_2 + \beta M)a \tag{5.55}$$

or:

$$a = \frac{(m_2 - m_1)g}{(m_1 + m_2 + \beta M)} \tag{5.56}$$

This is almost like the previous solution – and indeed, in the limit $M \to 0$ ***is*** the previous solution – but the net force between the two masses now must also *partially* accelerate the mass of the pulley. Partially because only the mass near the rim of the pulley is accelerated at the full rate a – most of the mass near the middle of the pulley has a much lower acceleration.

Note also that if $M = 0$, $T_1 = T_2$! This justifies – very much after the fact – our assertion early on that for a massless pulley, the tension in the string is everywhere constant. Here we see *why* that is true – because in order for the tension in the string between two points to be different, there has to be some *mass* in between those points for the force difference to act upon! In this problem, that mass is *the pulley*, and to keep the pulley accelerating up with the string, the string has to exert a *torque* on the pulley due to the unequal forces.

One last thing to note. We are being rather cavalier about the *normal* force exerted by the pulley on the string – all we can easily tell is that the total force acting up on the string must equal $T_1 + T_2$, the force that the string pulls down on the pulley with. Similarly, since the center of mass of the pulley does not move, we have something like $T_p - T_1 - T_2 - Mg = 0$. In other words, there are other questions I could always ask about pictures like this one, and by now you should have a good idea how to answer them.

5.7: Rotational Work and Energy

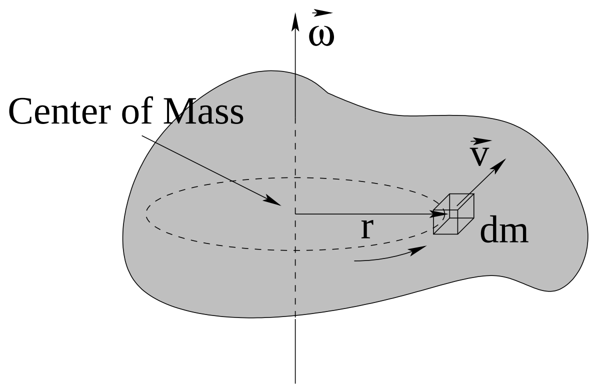


Figure 5.13: A blob of mass rotates about an axis through the center of mass, with an angular velocity as shown.

We have already laid the groundwork for studying work and energy in rotating systems. Let us consider the kinetic energy of an object rotating around its center of mass as portrayed in figure 5.13. The center of mass is at rest in this figure, so this is a *center of mass inertial coordinate system*.

It is easy for us to write down the kinetic energy of the little chunk of mass dm drawn into the figure at a distance r from the axis of rotation. It is just:

$$dK_{\rm in\ CM} = \frac{1}{2} dm v^2 = \frac{1}{2} dm r^2 \Omega^2 \tag{5.57}$$

To find the total, we integrate over all of the mass of the blob:

$$K_{\rm in\ CM} = \frac{1}{2} \int_{\rm blob} dm r^2 \Omega^2 = \frac{1}{2} I \Omega^2 \tag{5.58}$$

which works because Ω ***is the same for all chunks dm in the blob*** and is hence a constant that can be taken out of the integral, leaving us with the integral for I.

If we combine this with the theorem proved at the end of the last chapter we at last can precisely describe the kinetic energy of a rotating baseball in rest frame of the ground:

$$K = \frac{1}{2} M v_{\rm cm}^2 + \frac{1}{2} I \Omega^2 \tag{5.59}$$

That is, the kinetic energy in the lab is the kinetic energy ***of*** the (mass moving as if it is all at the) center of mass plus the kinetic energy ***in*** the center of mass frame, $\frac{1}{2} I \Omega^2$. We'll have a bit more to say about this when we prove the parallel axis theorem later.

5.7.1: Work Done on a Rigid Object

We have already done rotational work. Indeed we *began* with rotational work in order to obtain Newton's Second Law for one dimensional rotations above! However, there is much

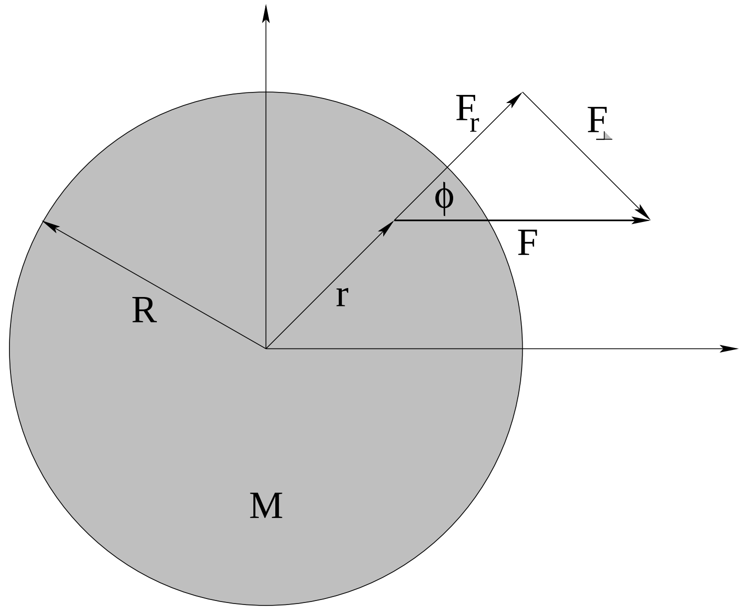


Figure 5.14: A force $\vec{F}$ is applied in an arbitrary direction at an arbitrary point on an arbitrary rigid object, decomposed in a center of mass coordinate frame. A disk is portrayed only because it makes it easy to see where the center of mass is.

to be gained by considering the *total* work done by an *arbitrary* force acting on an arbitrary extended rigid mass. Consider the force $\vec{F}$ in figure **??**, where I drew a regular shape (a disk) only to make it easy to see and draw an useful center of mass frame – it could just as easy be a force applied to the blob-shaped mass above in figure **??**.

I have decomposed the force $\vec{F}$ into two components:

$$F_r = F\cos(\phi) \tag{5.60}$$
$$F_\perp = F\sin(\phi) \tag{5.61}$$
$$\tag{5.62}$$

Suppose that this force acts for a short time dt, beginning (for convenience) with the mass at rest. We expect that the work done will consist of two parts:

$$dW = F_r dr + F_\perp ds \tag{5.63}$$

where $ds = r\ d\theta$. This is then:

$$dW = W_r + W_\theta = F_r dr + rF_\perp d\theta = F_r dr + \tau\ d\theta \tag{5.64}$$

We know that:

$$dW_r = F_r\ dr = dK_r \tag{5.65}$$
$$dW_\theta = \tau\ d\theta = dK_\theta \tag{5.66}$$

and if we integrate these independently, we get:

$$W_{\rm tot} = W_{\rm cm} + W_\theta = \Delta K_{\rm cm} + \Delta K_\theta \tag{5.67}$$

or the ***work decomposes into two parts!*** The work done by the component of the force *through* the center of mass accelerates the center of mass and changes the kinetic energy ***of*** the center of mass of the system as if it is a particle! The work done by the component of the force ***perpendicular*** to the line connecting the center of mass to the point where the force is applied to the rigid object increases the ***rotational kinetic energy***, the kinetic energy ***in*** the center of mass frame.

Hopefully this is all making a certain amount of rather amazing, terrifying, *sense* to you. One reason that torque and rotational physics is so important is that we can cleanly decompose the physics of rotating rigid objects *consistently, everywhere* into the physics of the motion of the center of mass and *rotation* about the center of mass. Note well that we have also written the WKE theorem in rotational terms, and are now justified in using *all* of the results of the work and energy chapter/week in (fully or partially) rotational problems!

Before we start, though, let's think a teeny bit about the rolling constraint and work, as we will be solving many rolling problems.

5.7.2: The Rolling Constraint and Work

A car is speeding down the highway at 50 meters per second (quite fast!) being chased by the police. Its tires hum as they *roll* down the highway without sliding. Fast as it is going, there are *four points* on this car that are *not moving at all relative to the ground!* Where are they?

The four places where the tires are in contact with the pavement, of course. Those points aren't *sliding* on the pavement, they are rolling, and "rolling" means that they are coming down at rest onto the pavement and then lifting up again as the tire rolls on.

If the car is travelling at a constant speed (and we neglect or arrange for there to be no drag/friction) we expect that the road will exert *no* force along the direction of motion of the car – the force exerted by static friction will be zero. Indeed, that's why wheels were invented – an object that is rolling at constant speed on frictionless bearings requires no force to keep it going – wheels are a way to avoid kinetic/sliding friction altogether!

More reasonably, the force exerted by static friction will *not* be zero, though, when the car speeds up, slows down, climbs or descends a hill, goes around a banked turn, overcomes drag forces to maintain a constant speed.

What happens to the *energy* in all of these cases, when the only force exerted by the ground is static friction at the points where the tire touches the ground? What is the work done by the force of static friction acting on the tires?

Zero! The force of static friction does *no* work on the system.

If you think about this for a moment, this result is almost certain to make your head ache. On the one hand, it is obvious:

$$dW_s = F_s dx = 0 \tag{5.68}$$

because $dx = 0$ in the frame of the ground – the place where the tires touch the ground *does not move*, so the force of static friction acts through *zero distance* and **does no work**.

Um, but if static friction does no work, how does the car speed up (you might ask)? What *else* could be doing work on the car? Oooo, head-starting-to-huuuurrrrrt...

Maybe, I dunno, *the motor?*

In fact, the car's engine exerts a *torque* on the wheels that is opposed by the *pivot force* at the road – the point of contact of a rolling object is a *natural pivot* to use in a problem, because forces exerted there, in addition to doing no work, exert no torque about that particular pivot. By fixing that pivot point, the car's engine creates a net torque that accelerates the wheels and, since they are fixed at the pivot, propels the car forward. Note well that the actual source of energy, however, is the *engine*, not the *ground*. This is key.

In general, in work/energy problems below, we will treat the force of static friction in rolling problems (disks, wheels, tires, pulleys) where there is rolling without slipping as ***doing no work*** and hence acting like a normal force or other force of constraint – not exactly a "conservative force" but one that we can ignore when considering the Generalized Non-Conservative Work – Mechanical Energy theorem or just the plain old WKE theorem solving problems.

Later, when we consider pivots in collisions, we will see that pivot forces *often* cause momentum not to be conserved – another way of saying that they can cause energy to enter or leave a system – but that they are generally not the *source* of the energy. Like a skilled martial artist, they do not provide energy themselves, but they are very effective at diverting energy from one form to another. In fact, *very* much like a skilled martial artist, come to think about it.

It isn't really a metaphor...

Example 5.7.1: Work and Energy in Atwood's Machine

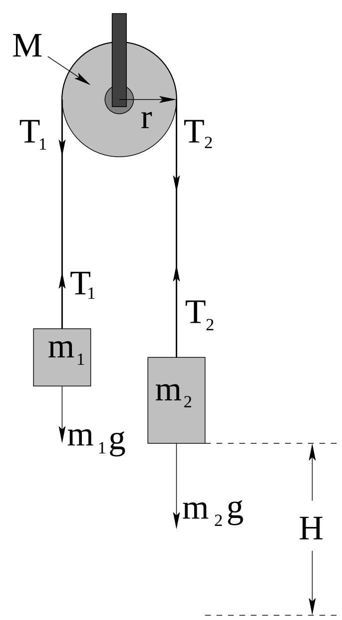


Figure 5.15: Atwood's Machine, but this time with a *massive* pulley of mass M and radius R (and moment of inertia $I = \beta M R^2$), this time solving a "standard" conservation of mechanical energy problem.

We would like to find the speed v of m_1 and m_2 (and the angular speed Ω of mass M) when mass $m_2 > m_1$ falls a height H, beginning from rest, when the massless unstretchable string connecting the masses rolls without slipping on the massive pulley. We *could* do this problem using a from the solution to the example above, finding the time t it takes to reach H, and backsubstituting to find v, but by now we know quite well that it is a lot easier to use energy conservation (since no non-conservative forces act if the string does not slip) which is already time-independent.

Figure 5.15 shows the geometry of the problem. Note well that mass m_1 will go *up* a distance H at the same time m_2 goes *down* a distance H.

Again our solution strategy is almost identical to that of the conservation of mechanical energy problems of two weeks ago. We simply evaluate the initial and final total mechanical energy ***including the kinetic energy of the pulley*** and ***using the rolling constraint*** and solve for v. We can choose the zero of potential energy for the two mass separately, and choose to start m_2 a height H above its final position, and we start mass m_1 at zero potential. The final potential energy of m_2 will thus be zero and the final potential energy of m_1 will be $m_1 gH$. Also, we will need to substitute the rolling constraint into the expression for the rotational kinetic energy of the pulley in the little patch of algebra below:

$$\Omega = \frac{v}{r} \tag{5.69}$$

Thus:

$$\begin{aligned}
E_i &= m_2 gH \\
E_f &= m_1 gH + \frac{1}{2}m_1 v^2 + \frac{1}{2}m_2 v^2 + \frac{1}{2}\beta M R^2 \Omega^2 \qquad \text{or} \\
m_2 gH &= m_1 gH + \frac{1}{2}m_1 v^2 + \frac{1}{2}m_2 v^2 + \frac{1}{2}\beta M R^2 \Omega^2
\end{aligned}$$

and now we substitute the rolling constraint:

$$\begin{aligned}
(m_2 - m_1)gH &= \frac{1}{2}(m_1 + m_2)v^2 + \frac{1}{2}\beta M R^2 \frac{v^2}{R^2} \\
(m_2 - m_1)gH &= \frac{1}{2}(m_1 + m_2 + \beta M)v^2 \qquad (5.70)
\end{aligned}$$

to arrive at

$$v = \sqrt{\frac{2(m_2 - m_1)gH}{m_1 + m_2 + \beta M}} \tag{5.71}$$

You can do this! It isn't really that difficult (or that *different* from what you've done before).

Note well that the pulley *behaves like an extra mass* βM in the system – all of this mass has to be accelerated by the actual force difference between the two masses. If $\beta = 1$ – a ring of mass – then all of the mass of the pulley ends up moving at v and *all* of its mass counts. However, for a disk or ball or actual pulley, $\beta < 1$ because *some* of the rotating pulley's mass is moving more slowly than v and has less kinetic energy when the pulley is rolling.

Also note well that the strings do no *net* work in the system. They are internal forces, with T_2 doing negative work on m_2 but equal and opposite *positive* work on M, with T_1 doing negative work on M, but doing equal and opposite *positive* work on m_1. Ultimately, the tensions in the string serve only to *transfer energy* between the masses and the pulley so that the change in potential energy is correctly shared by all of the masses when the string rolls without slipping.

Example 5.7.2: Unrolling Spool

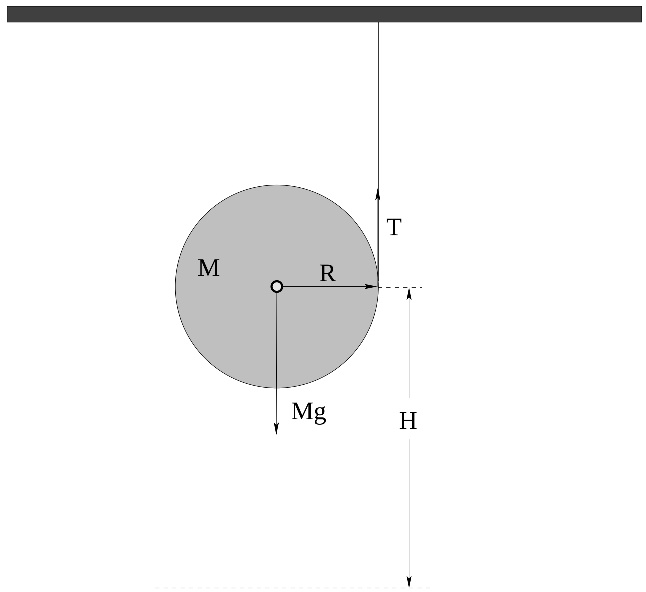


Figure 5.16: A spool of fishing line is tied to a pole and released from rest to fall a height H, unrolling as it falls.

In figure 5.16 a spool of fishing line that has a total mass M and a radius R and is effectively a disk is tied to a pole and released from rest to fall a height H. Let's find *everything*: the acceleration of the spool, the tension T in the fishing line, the speed with which it reaches H.

We start by writing Newton's Second Law for both the translational and rotational motion. We'll make down y-positive. Why not! First the force:

$$F_y = Mg - T = Ma \tag{5.72}$$

and then the torque:

$$\tau = RT = I\alpha = \left(\frac{1}{2}MR^2\right)\frac{a}{R} \tag{5.73}$$

We ***use the rolling constraint*** (as shown) to rewrite the second equation, and divide both sides by R. Writing the first and second equation together:

$$Mg - T = Ma \tag{5.74}$$

$$T = \frac{1}{2}Ma \tag{5.75}$$

we add them:

$$Mg = \frac{3}{2}Ma \tag{5.76}$$

and solve for a:

$$a = \frac{2}{3}g \tag{5.77}$$

We back substitute to find T:

$$T = \frac{1}{3}Mg \tag{5.78}$$

Next, we tackle the energy conservation problem. I'll do it really fast and easy:

$$E_i = MgH = \frac{1}{2}Mv^2 + \frac{1}{2}\left(\frac{1}{2}MR^2\right)\left(\frac{v}{R}\right)^2 = E_f \tag{5.79}$$

or

$$MgH = \frac{3}{4}Mv^2 \tag{5.80}$$

and

$$v = \sqrt{\frac{4gH}{3}} \tag{5.81}$$

Example 5.7.3: A Rolling Ball Loops-the-Loop

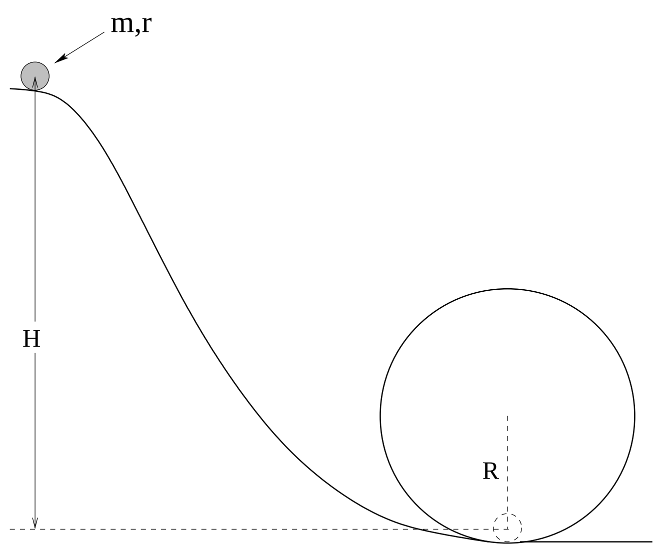


Figure 5.17: A ball of mass m and radius r rolls without slipping to loop the loop on the circular track of radius R.

Let's redo the "Loop-the-Loop" problem, but this time let's consider a ***solid ball*** of mass m and radius r going around the track of radius R. This is a tricky problem to do *precisely* because as the normal force decreases (as the ball goes around the track) at *some* point the static frictional force required to "keep the ball rolling" on the track may well become greater than $\mu_s N$, at which point the ball will *slip*. Slipping dissipates energy, so one would have to raise the ball slightly at the beginning to accommodate this. Of course, raising the ball slightly at the beginning also increases N so maybe it doesn't ever slip. So the best we can solve for is the minimum height $H_{\rm min}$ it must have to roll without slipping assuming that it doesn't ever actually slip, and "reality" is probably a bit higher to accommodate or prevent slipping, overcome drag forces, and so on.

With that said, the problem's solution is *exactly the same as before* except that in the energy conservation step one has to use:

$$mgH_{\rm min} = 2mgR + \frac{1}{2}mv_{\rm min}^2 + \frac{1}{2}I\Omega_{\rm min}^2 \tag{5.82}$$

plus the rolling constraint $\Omega_{\rm min} = v_{\rm min}/r$ to get:

$$\begin{aligned} mgH_{\rm min} &= 2mgR + \frac{1}{2}mv_{\rm min}^2 + \frac{1}{5}mv_{\rm min}^2 \\ &= 2mgR + \frac{7}{10}mv_{\rm min}^2 \end{aligned} \quad (5.83)$$

Combine this with the usual:

$$mg = \frac{mv_{\rm min}^2}{R} \quad (5.84)$$

so that the ball "barely" loops the loop and you get:

$$H_{\rm min} = 2.7R \quad (5.85)$$

only a tiny bit higher than needed for a block sliding on a frictionless track.

Really, not all that difficult, right? All it takes is some *practice*, both redoing these examples on your own and doing the homework and it will all make sense.

5.8: The Parallel Axis Theorem

As we have seen, the moment of inertia of an object or collection of point-like objects is just

$$I = \sum_i m_i r_i^2 \quad (5.86)$$

where r_i is the distance between the axis of rotation and the point mass m_i in a rigid system, or

$$I = \int r^2 dm \quad (5.87)$$

where r is the distance from the axis of rotation to "point mass" dm in the rigid object composed of continuously distributed mass.

However, in the previous section, we saw that the kinetic energy of a rigid object relative to an *arbitrary* origin can be written as the sum of the kinetic energy of the object itself treated as a total mass located at the (moving) center of mass plus the kinetic energy of the object in the moving center of mass reference frame.

For the particular case where a rigid object rotates *uniformly* around an axis that is parallel to an axis through the center of mass of the object, that is, in such a way that the angular velocity *of* the center of mass equals the angular velocity *around* the center of mass we can derive a theorem, called the ***Parallel Axis Theorem***, that can greatly simplify problem solving while embodying the previous result for the kinetic energies. Let's see how.

Suppose we want to find the moment of inertia of the arbitrary "blob shaped" rigid mass distribution pictured above in figure **??** about the axis labelled "New (Parallel) Axis". This is, by

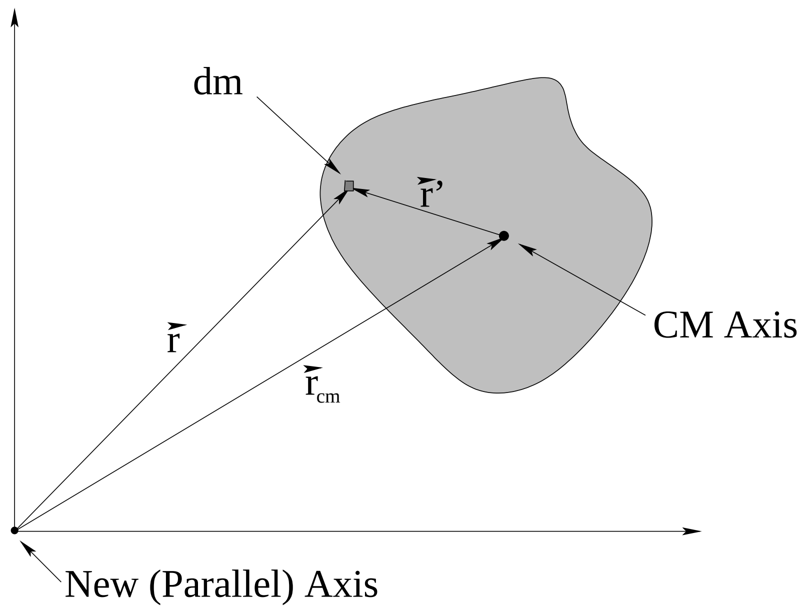


Figure 5.18: An arbitrary blob of total mass M rotates around the axis at the origin as shown. Note well the geometry of $\vec{r}_{\rm cm}$, $\vec{r}'$, and $\vec{r} = \vec{r}_{\rm cm} + \vec{r}'$.

definition (and using the fact that $\vec{r} = \vec{r}_{\rm cm} + \vec{r}'$ from the triangle of vectors shown in the figure):

$$\begin{aligned}
I &= \int r^2\, dm \\
&= \int \left(\vec{r}_{\rm cm} + \vec{r}'\right) \cdot \left(\vec{r}_{\rm cm} + \vec{r}'\right)\, dm \\
&= \int \left(r_{\rm cm}^2 + r'^2 + 2\vec{r}_{\rm cm} \cdot \vec{r}'\right)\, dm \\
&= \int r_{\rm cm}^2\, dm + \int r'^2\, dm + 2\vec{r}_{\rm cm} \cdot \int \vec{r}'\, dm \\
&= r_{\rm cm}^2 \int dm + \int r'^2\, dm + 2M\vec{r}_{\rm cm} \cdot \left(\frac{1}{M}\int \vec{r}'\, dm\right) \\
&= M r_{\rm cm}^2 + I_{\rm cm} + 2M\vec{r}_{\rm cm} \cdot (0)
\end{aligned} \tag{5.88}$$

or

$$I = I_{\rm cm} + M r_{\rm cm}^2 \tag{5.89}$$

In case that was a little fast for you, here's what I did. I substituted $\vec{r}_{\rm cm} + \vec{r}'$ for $\vec{r}$. I distributed out that product. I used the linearity of integration to write the integral of the sum as the sum of the integrals (all integrals over all of the mass of the rigid object, of course). I noted that $\vec{r}_{\rm cm}$ is a *constant* and pulled it out of the integral, leaving me with the integral $M = \int dm$. I noted that $\int r'^2 dm$ is just $I_{\rm cm}$, the moment of intertia of the object about an axis through its center of mass. I noted that $(1/M)\int \vec{r}' dm$ is the position *of* the center of mass in center of mass coordinates, which is *zero* – by definition the center of mass is at the origin of the center of mass frame.

The result, in words, is that the moment of inertia of an object that uniformly rotates around any axis is the moment of inertia of the object about an axis parallel to that axis through the center of mass of the object *plus* the moment of inertia of the total mass of the object treated as a point mass located at the center of mass as it revolves!

This sounds a lot like the kinetic energy theorem; let's see how the two are related.

As long as the object rotates uniformly – that is, the object goes around its own center one time for every time it goes around the axis of rotation, keeping the *same side pointing in towards the center as it goes* – then its kinetic energy is just:

$$K = \frac{1}{2} I \Omega^2 = \frac{1}{2}(M r_{\rm cm}^2)\Omega^2 + \frac{1}{2} I_{\rm cm}\Omega^2 \tag{5.90}$$

A bit of algebraic legerdemain:

$$K = \frac{1}{2}(M r_{\rm cm}^2)\left(\frac{v_{\rm cm}}{r_{\rm cm}}\right)^2 + \frac{1}{2} I_{\rm cm}\Omega^2 = K(\textit{of}\ {\rm cm}) + K(\textit{in}\ {\rm cm}) \tag{5.91}$$

as before!

Warning! This will not work if an object is revolving many times around its own center of mass for each time it revolves around the parallel axis.

Example 5.8.1: Moon Around Earth, Earth Around Sun

This is a conceptual example, not really algebraic. You may have observed that the Moon always keeps the same face towards the Earth – it is said to be "gravitationally locked" by tidal forces so that this is true. This means that the Moon revolves once on its axis in exactly the same amount of time that the Moon itself revolves around the Earth. We could therefore compute the *total* angular kinetic energy of the Moon by assuming that it is a solid ball of mass M, radius r, in an orbit around the Earth of radius R, and a period of 28.5 days:

$$I_{\rm moon} = MR^2 + \frac{2}{5}Mr^2 \tag{5.92}$$

(from the parallel axis theorem),

$$\Omega = \frac{2\pi}{T} \tag{5.93}$$

(you'll need to find T in seconds, 86400×28.5) and then:

$$K = \frac{1}{2} I_{\rm moon}\Omega^2 \tag{5.94}$$

All that's left is the arithmetic.

Contrast this with the Earth rotating around the Sun. *It* revolves on its own axis 365.25 times during the period in which it revolves around the Sun. To find *it's* kinetic energy we could *not* use the parallel axis theorem, but we can still use the theorem at the end of the previous chapter. Here we would find two different angular velocities:

$$\Omega_{\rm day} = \frac{2\pi}{T_{\rm day}} \tag{5.95}$$

and

$$\Omega_{\rm year} = \frac{2\pi}{T_{\rm year}} \tag{5.96}$$

(again, 1 day = 86400 seconds is a good number to remember). Then if we let M be the mass of the Earth, r be its radius, and R be the radius of its orbit around the Sun (all numbers that are readily available on Wikipedia[134] we could find the total kinetic energy (relative to the Sun) as:

$$K = \frac{1}{2}\left(MR^2\right)\Omega^2_{\rm year} + \frac{1}{2}\left(\frac{2}{5}Mr^2\right)\Omega^2_{\rm day} \tag{5.97}$$

which is somewhat more complicated, no?

Let's do a more readily evaluable example:

Example 5.8.2: Moment of Inertia of a Hoop Pivoted on One Side

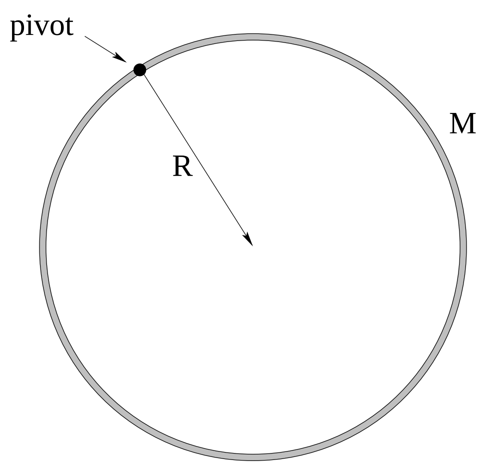


Figure 5.19: A hoop of mass M and radius R is pivoted on the side – think of it as being hung on a nail from a barn door.

In figure 5.19 a hoop of mass M and radius R is pivoted at a point on the *side*, on the hoop itself, not in the middle. We already know the moment of inertia of the hoop about its center of mass. What is the moment of inertia of the hoop about this new axis *parallel* to the one through the center of mass that we used before?

It's so simple:

$$I_{\rm side\ pivot} = MR^2 + I_{\rm cm} = MR^2 + MR^2 = 2MR^2 \tag{5.98}$$

and we're done!

For your homework, you get to evaluate the moment of inertia of a rod about an axis through its center of mass and about one end of the rod and compare the two, both using direct integration and using the parallel axis theorem. Good luck!

5.9: Perpendicular Axis Theorem

In the last section we saw how a bit of geometry and math allowed us to prove a very useful theorem – useful because we can now learn a short table of moments of inertia about a given

[134] Wikipedia: http://www.wikipedia.org/wiki/Earth.

axis through the center of mass and then *easily extend them* to find the moments of inertia of these same shapes when they uniformly rotate around a parallel axis.

In this section we will similarly derive a theorem that is very useful for relating moments of inertia of ***planar distributions of mass*** (only) around axes that are *perpendicular* to one another – the ***Perpendicular Axis Theorem***. Here's how it goes.

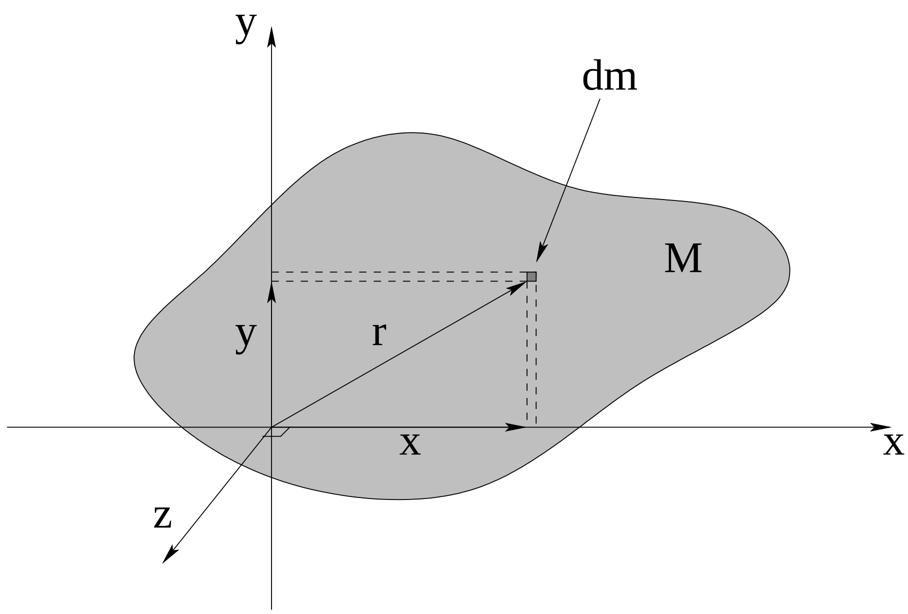


Figure 5.20: A planar blob of mass and the geometry needed to prove the Perpendicular Axis Theorem

Suppose that we wish to evaluate I_x, the moment of inertia of the plane mass distribution M shown in figure 5.20. That's quite easy:

$$I_x = \int y^2 dm \tag{5.99}$$

Similarly,

$$I_y = \int x^2 dm \tag{5.100}$$

We add them, and presto chango!

$$I_x + I_y = \int y^2 dm + \int x^2 dm = \int r^2 dm = I_z \tag{5.101}$$

This is it, the ***Perpendicular Axis Theorem:***

$$I_z = I_x + I_y \tag{5.102}$$

I'll give a single example of its use. Let's find the moment of inertia of a hoop about an axis through the center ***in the plane of the loop!***

Example 5.9.1: Moment of Inertia of Hoop for Planar Axis

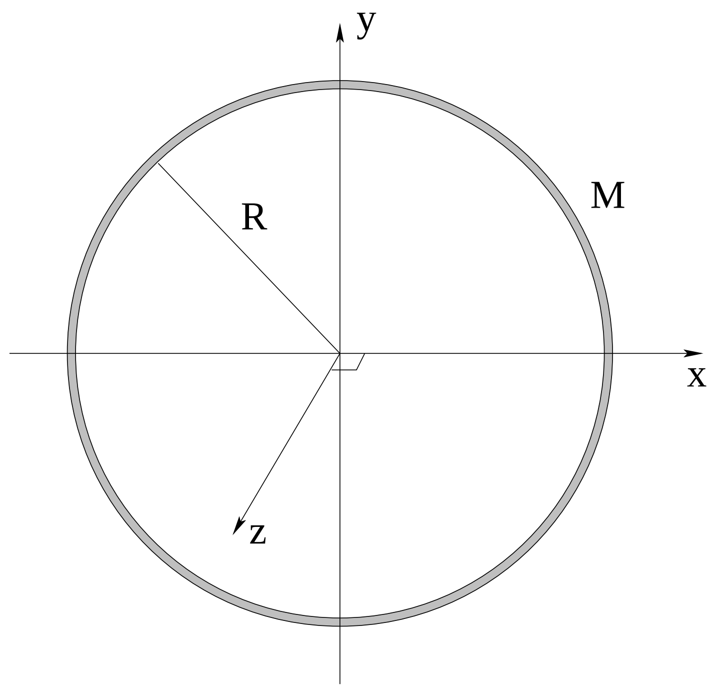


Figure 5.21: A hoop of mass M and radius R is drawn. What is the moment of inertia about the x-axis?

This one is really very, very easy. We use the Perpendicular Axis theorem *backwards* to get the answer. In this case we know $I_z = MR^2$, and want to find I_x. We observe that *from symmetry*, $I_x = I_y$ so that:

$$I_z = MR^2 = I_x + I_y = 2I_x \tag{5.103}$$

or

$$I_x = \frac{1}{2} I_z = \frac{1}{2} MR^2 \tag{5.104}$$

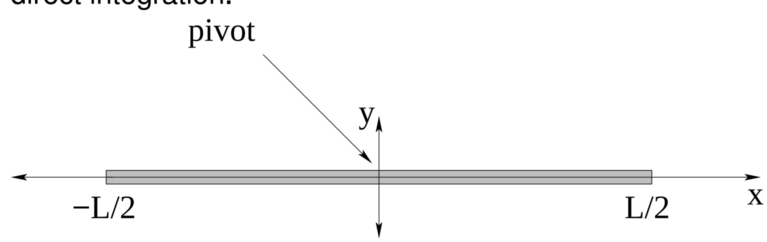
Homework for Week 5

Problem 1.

Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

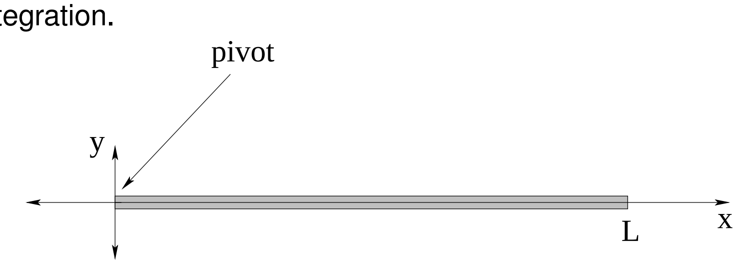
Problem 2.

a) Evaluate the moment of inertia of a uniform rod of mass M and length L about its center of mass by direct integration.



This is one of the moments you should "just know".

b) Evaluate the moment of inertia of a uniform rod of mass M and length L about one end by direct integration.

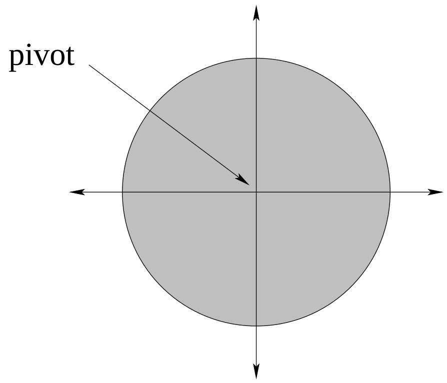


This too is a moment of inertia you should just know.

c) ***Also*** evaluate the moment of inertia about one end using the parallel axis theorem and the result you just obtained in a). Which (if you know $I_{\rm cm}$) is easier?

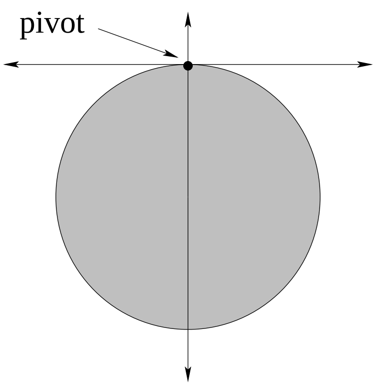
Problem 3.

a)



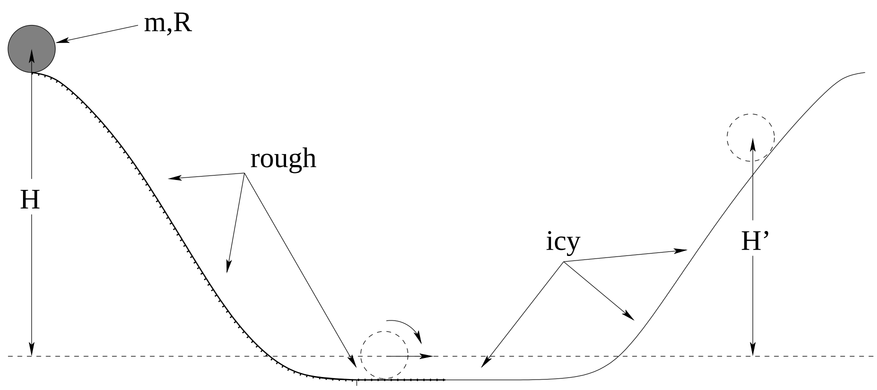
Evaluate the moment of inertia of a uniform disk of mass M and radius R about its axis of symmetry by direct integration (this can be set up as a "one dimensional integral" and hence is not too difficult).

b)



Evaluate the moment of inertia of this disk around a pivot at the edge of the disk (for example, a thin nail stuck through a hole at the outer edge) ***using the parallel axis theorem***. Would you care to do the actual integral to find the moment of inertia of the disk in this case?

Problem 4.



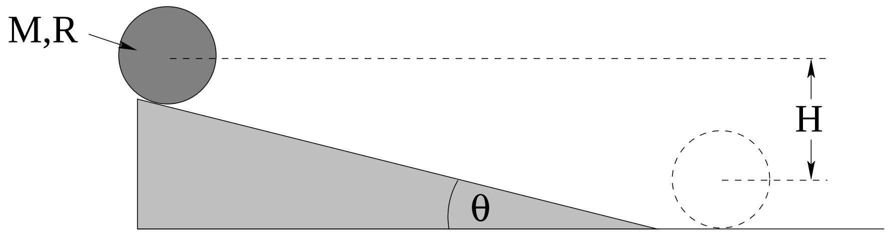
A disk of mass m and radius R ***rolls without slipping*** down a rough slope of height H onto an icy (frictionless) track at the bottom that leads up a second ***icy (frictionless)*** hill as shown.

a) How fast is the disk moving at the bottom of the first incline? How fast is it rotating (what is its angular velocity Ω)?

b) Does the disk's angular velocity change as it leaves the rough track and moves onto the ice (in the middle of the flat stretch in between the hills)?

c) How far up the second hill (vertically, find H') does the disk go before it stops rising?

Problem 5.



A round object with mass m and radius R is released from rest to roll ***without slipping*** down an inclined plane of height H at angle θ relative to horizontal. The object has a moment of inertia $I = \beta m R^2$ (where β is a dimensionless number such as $\frac{1}{2}$ or $\frac{2}{5}$, that might describe a disk or a solid ball, respectively).

a) Begin by putting down the rolling constraint: relating v (the speed of the center of mass) to the angular velocity (for the rolling object). You will use this (and the related two equations for s and θ and a and α) repeatedly in rolling problems.

b) Using Newton's second law in both its linear and rotational form plus the rolling constraint, show that the acceleration of the object is:

$$a = \frac{g \sin\theta}{1 + \beta}$$

c) Using conservation of mechanical energy, show that it arrives at the bottom of the incline with a velocity:

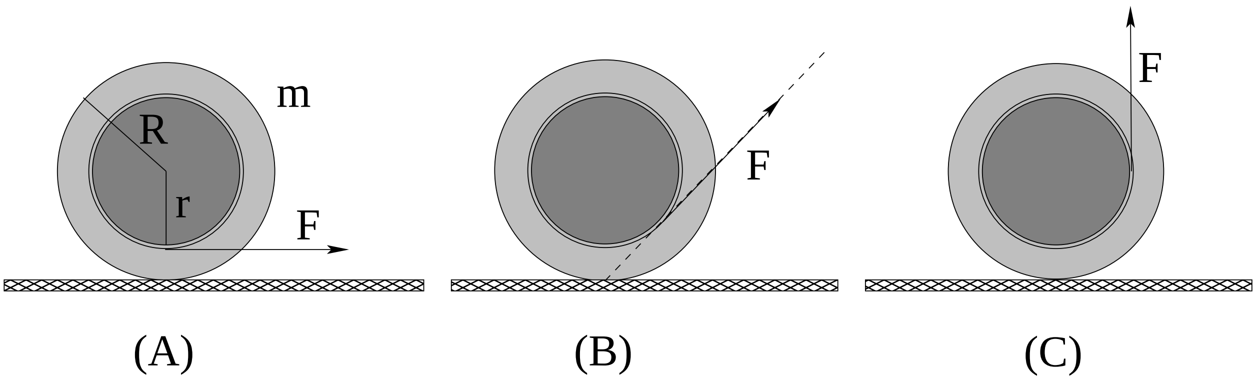
$$v = \sqrt{\frac{2gH}{1 + \beta}}$$

d) Show that the condition for the greatest angle for which the object will roll *without slipping* is that:

$$\theta < \theta_c = \tan^{-1}\left(\mu_s \frac{1 + \beta}{\beta}\right)$$

where μ_s is the coefficient of static friction between the object and the incline. Remember! $f_s < \mu_s N$! Not equal to!

Problem 6.



In the figure above, a spool of mass m is wrapped with string around the inner spool. The spool is placed on a rough surface (with coefficient of friction $\mu_s = 0.5$) and the string is pulled with force $F \leq \frac{1}{4}mg$ in the three directions shown.

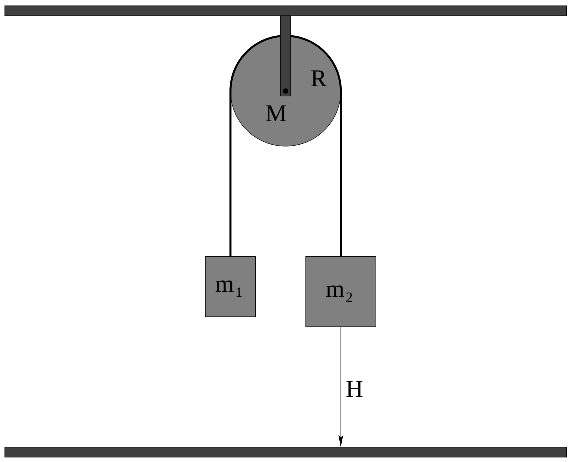
a) For each picture, indicate the direction that static friction will point. Can the spool slip while it rolls for this magnitude of force?

b) For each picture, indicate the direction that the spool will accelerate.

c) For each picture, find the magnitude of the force exerted by static friction and the magnitude of the acceleration of the spool in terms of r, R or $I_{\rm cm} = \beta m R^2$.

Note: You can use ***either*** the center of mass ***or*** the point of contact with the ground (with the parallel axis theorem) as a pivot, the latter being ***slightly easier*** both algebraically and intuitively.

Problem 7.



In the figure provided, Atwood's machine is drawn – two masses m_1 and m_2 hanging over a ***massive*** pulley which you can model as a disk of mass M and radius R, connected by a massless unstretchable string. The string rolls on the pulley without slipping and there is no friction in the bearing.

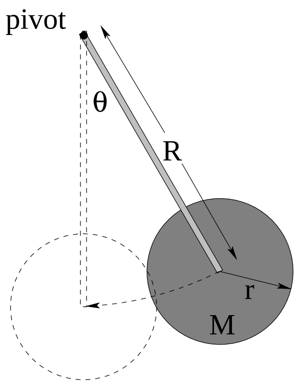
a) Draw free body diagrams or force diagrams directly onto (a copy of) the picture, showing both the forces and *where* the forces act on each of the three masses.

b) Convert your diagram for each object into a statement of Newton's Second Law(s) either linear or rotational as appropriate – for that object.

c) Using the ***rolling constraint*** for the pulley, find the magnitude of the acceleration of either mass, the tension(s) in the string on *both* sides of the pulley in terms of m_1, m_2, M, g, and R, and α, the angular acceleration of the pulley.

d) Suppose mass $m_2 > m_1$ and the system is released from rest with the masses at equal heights. When mass m_2 has descended a distance H, ***use conservation of mechanical energy*** to find velocity of each mass and the angular velocity of the pulley.

Problem 8.



In the picture to the left, a physical pendulum is constructed by hanging a disk of mass M and radius r on the end of a massless rigid rod in such a way that the center of mass of the disk is a distance R away from the pivot and so that the whole disk pivots with the rod. The pendulum is pulled to an initial angle θ_0 (relative to vertically down) and then released.

a) Find the *torque* about the pivot exerted on the pendulum by gravity at an arbitrary angle θ. Make positive θ ***out of the page*** so the angle illustrated above is *positive*.

b) Integrate the torque from $\theta = \theta_0$ to $\theta = 0$ to find the total work done by the gravitational torque as the pendulum disk falls to its lowest point. Note that your answer should be $MgR(1 - \cos(\theta_0)) = MgH$ where H is the initial height above this lowest point.

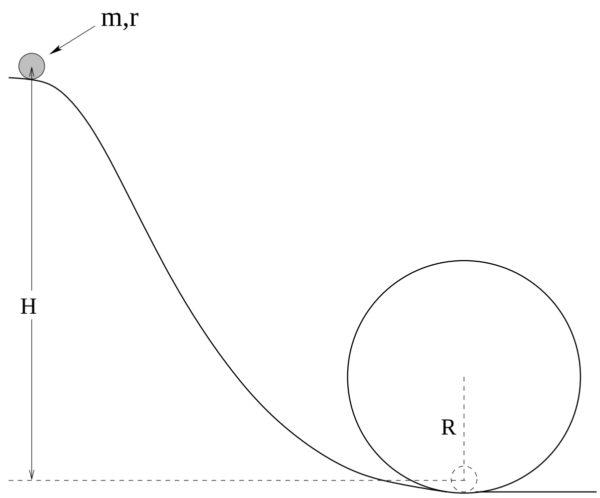
c) Find the moment of inertia of the pendulum about the pivot (using the parallel axis theorem).

d) Set the work you evaluated in b) equal to the rotational kinetic energy of the disk $\frac{1}{2}I\Omega^2$ using the moment of inertia you found in c). Solve for $\Omega = \frac{d\theta}{dt}$ when the disk is at its lowest point.

e) Show that this kinetic energy is equal to the kinetic energy of the moving center of mass of the disk $\frac{1}{2}Mv^2$ plus the kinetic energy of the disk's rotation about its own center of mass, $\frac{1}{2}I_{\rm cm}\omega^2$, at the lowest point.

In a few weeks we will learn to solve N2 for this problem for small angle oscillations, so it is good to work through it now.

Problem 9.



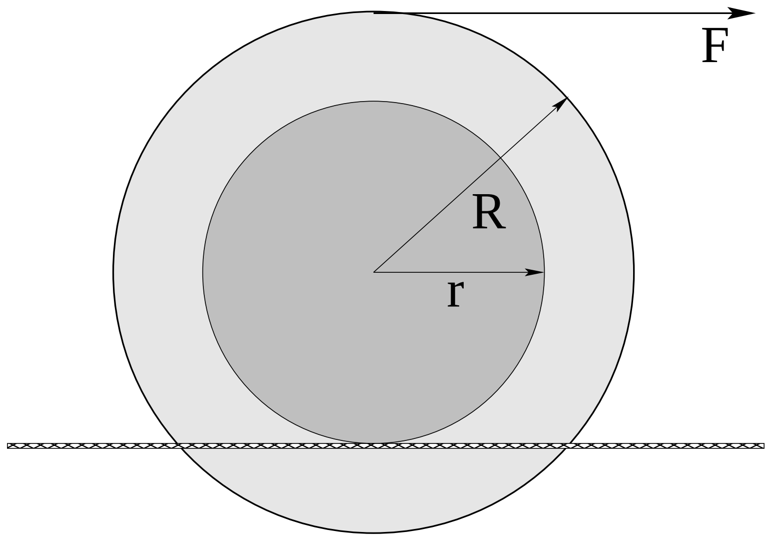
A solid ball of mass M and radius r sits at rest at the top of a hill of height H leading to a circular loop-the-loop. The center of mass of the ball will move in a circle of radius R if it goes around the loop. Recall that the moment of inertia of a solid ball is $I_{\rm ball} = \frac{2}{5}MR^2$.

a) Find the minimum height $H_{\rm min}$ for which the ball *barely* goes around the loop staying on the track at the top, ***assuming that it rolls without slipping***[135] the entire time independent of the normal force.

b) How does your answer relate to the minimum height for the earlier homework problem where it was a block that slid around a frictionless track? Does this answer make sense? If it is higher, where did the extra potential energy go? If it is lower, where did the extra kinetic energy come from?

[135]Something that will mostly likely not be true as $N \to 0$, although figuring out whether or not it slips a bit at the very top or *where* (at what angle) it slips is a ***hard problem!*** Too difficult to even be just "advanced" as far as this textbook is concerned...

Problem 10.



A cable spool of mass M, radius R and moment of inertia $I = \beta M R^2$ is wrapped around its OUTER disk with fishing line and the inner spindle of radius r is placed on a rough rope as shown. The fishing line is then pulled with a force F to the right so that the spool rolls down the rope to the right ***without slipping***.

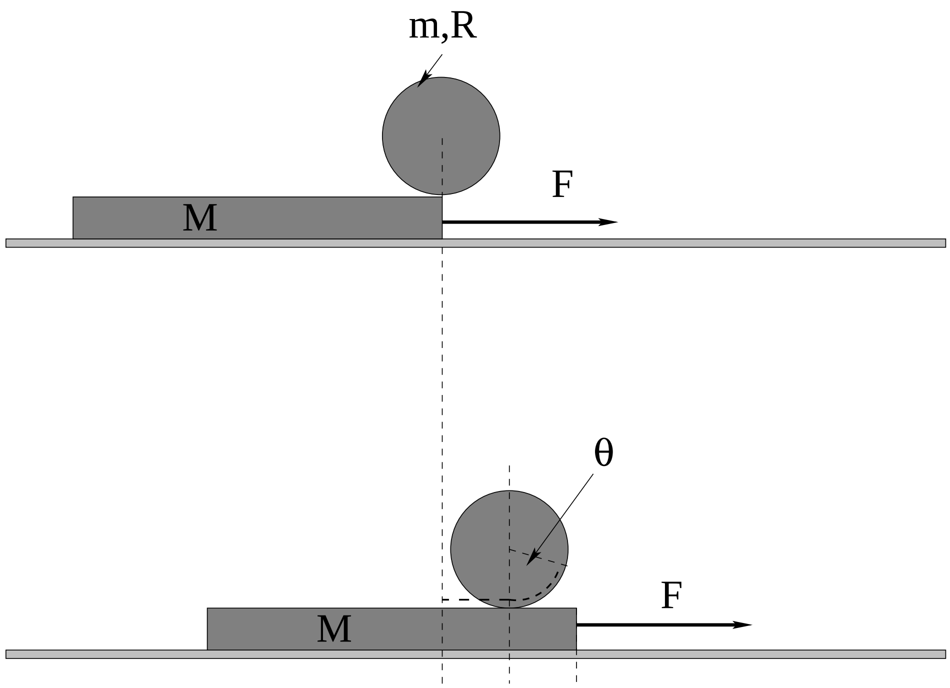
a) Find the magnitude and direction of the acceleration of the spool.

b) Find the force (magnitude and direction) exerted by the friction of the rope on the spool.

c) For one particular value of r, the magnitude of frictional force is *zero!*. Find that value (in terms of e.g. β and R). For larger values of r which way does friction point? For smaller values of r which way does friction point?

In analyzing the "walking the spool" problem in class and in the text above, students often ask how they can predict which direction that static friction acts on a rolling spool, and I reply that *they can't!* I can't, not always, because in this problem it can point *either way* and which way it ultimately points depends on the *details* of R, r and β! The best you can do is make a reasonable *guess* as to the direction and let the algebra speak – if your answer for your initial choice comes out negative, *friction points the other way*.

Advanced Problem 11.



A disk of mass m is resting on a slab of mass M, which in turn is resting on a frictionless table. The coefficients of static and kinetic friction between the disk and the slab are μ_s and μ_k, respectively. A small force $\vec{F}$ to the right is applied to the slab as shown, then gradually increased.

a) When F (the magnitude of $\vec{F}$) is small, the slab will accelerate to the right and the disk will roll on the slab without slipping. Find the acceleration of the slab a_s, the acceleration of the disk a_d, and the *angular* acceleration of the disk α as this happens

b) The disk ***will not slip if $F < F_{\mathbf{max}}$***. Find $F_{\max}$.

c) Suppose $F \geq F_{\max}$ (so the disk will definitely slip). Solve once again for a'_s, a'_d and α' (note primes).

Hint: The hardest single thing about this problem isn't the physics (which is really pretty straightfoward). It is visualizing the coordinates as the center of mass of the disk moves with a different acceleration as the slab. I have drawn *two* figures above to help you with this – the lower figure represents a possible position of the disk after the slab has moved some distance to the right and the disk has rolled back (Relative to the slab! It has moved *forward* relative to the ground! Why?) ***without slipping***. Note the dashed radius to help you see the angle through which it has rolled and the various dashed lines to help you relate the distance the slab has moved x_s, the distance the center of the disk has moved x_d, and the angle through which it has rolled θ. Use this relation to connect the acceleration of the slab to the acceleration and angular acceleration of the disk.

Week 6: Vector Torque and Angular Momentum

1.12: Summary

- The vector torque acting on a point particle or rigid body is:

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{r}} \times \vec{\boldsymbol{F}}$$

where $\vec{\boldsymbol{r}}$ is the vector from the ***pivot point*** (not axis!) to the point where the force is applied.

- The vector angular momentum of a point particle is:

$$\vec{\boldsymbol{L}} = \vec{\boldsymbol{r}} \times \vec{\boldsymbol{p}} = m(\vec{\boldsymbol{r}} \times \vec{\boldsymbol{v}})$$

where as before, $\vec{\boldsymbol{r}}$ is a vector from the ***pivot point*** to the location of the particle and $\vec{\boldsymbol{v}}$ is the particle's velocity.

- The vector form for Newton's Second Law for Rotation for a point particle is:

$$\vec{\boldsymbol{\tau}} = \frac{d\vec{\boldsymbol{L}}}{dt}$$

- All of these relations generalize when computing the ***total vector torque*** acting on a collection of particles (that may or may not form a rigid body) with a ***total angular momentum.*** Provided that all the internal forces $\vec{\boldsymbol{F}}_{ij} = -\vec{\boldsymbol{F}}_{ji}$ act along the lines $\vec{\boldsymbol{r}}_{ij}$ connecting the particles, there is no net torque due to the internal forces between particles and we get the series of results:

$$\vec{\boldsymbol{\tau}}_{\rm tot} = \sum_i \vec{\boldsymbol{\tau}}_i^{\rm ext}$$

$$\vec{\boldsymbol{L}}_{\rm tot} = \sum_i \vec{\boldsymbol{r}}_i \times \vec{\boldsymbol{p}}_i$$

and

$$\vec{\boldsymbol{\tau}}_{\rm tot} = \frac{d\vec{\boldsymbol{L}}_{\rm tot}}{dt}$$

- The ***Law of Conservation of Angular Momentum*** is:

  If (and only if) the total torque acting on a system is zero, then the total angular momentum of the system is a constant vector (conserved).

or in equationspeak:

If (and only if) $\vec{\tau}_{\rm tot} = 0$***, then*** $\vec{L}_{\rm tot}$ ***is a constant vector.***

- For rigid objects (or collections of point particles) that have ***mirror symmetry across the axis of rotation*** and/or ***mirror symmetry across the plane of rotation***, the vector angular momentum can be written in terms of the ***scalar moment of inertia*** about the axis of rotation (defined and used in week 5) and the ***vector angular velocity*** as:

$$\vec{L} = I\vec{\Omega}$$

- For rigid objects or collections of point particles that ***lack*** this symmetry with respect to an axis of rotation (direction of $\vec{\Omega}$)

$$\vec{L} \neq I\vec{\Omega}$$

  for any scalar I. In general, $\vec{L}$ *precesses* around the axis of rotation in these cases and requires a *constantly varying nonzero torque* to drive the precession.

- When two (or more) isolated objects collide, both momentum and angular momentum is conserved. Angular momentum conservation becomes an additional equation (set) that can be used in analyzing the collision.

- If one of the objects is *pivoted*, then angular momentum *about this pivot* is conserved but in general momentum is *not* conserved as the pivot itself will convey a significant impulse to the system during the collision.

- Radial forces – any force that can be written as $\vec{F} = F_r\vec{r}$ – ***exert no torque*** on the masses that they act on. Those object generally move in not-necessarily-circular ***orbits*** with ***constant angular momentum***.

- When a rapidly spinning symmetric rotator is acted on by a torque of constant magnitude that is (always) perpendicular to the plane formed by the angular momentum and a vector in second direction, the angular momentum vector ***precesses*** around the second vector. In particular, for a spinning top with angular momentum $\vec{L}$ tipped at an angle θ to the vertical, the magnitude of the torque exerted by gravity and the normal force on the top is:

$$\tau = |\vec{D} \times mg\hat{z}| = mgD\sin(\theta) = \left|\frac{d\vec{L}}{dt}\right| = L\sin(\theta)\Omega_p$$

  or

$$\Omega_p = \frac{mgD}{L}$$

  In this expression, Ω_p is the angular ***precession frequency*** of the top and $\vec{D}$ is the vector from the point where the tip of the top rests on the ground to the center of mass of the top. The direction of precession is determined by the ***right hand rule***.

6.1: Vector Torque

In the previous chapter/week we saw that we could describe rigid bodies rotating about a single axis quite accurately by means of a modified version of Newton's Second Law:

$$\tau = r_F F \sin(\phi) = |\vec{\boldsymbol{r}}_F \times \vec{\boldsymbol{F}}| = I\alpha \tag{6.1}$$

where I is the moment of inertia of the rigid body, evaluated by summing/integrating:

$$I = \sum_i m_i r_i^2 = \int r^2 dm \tag{6.2}$$

In the torque expression $\vec{\boldsymbol{r}}_F$ was a vector in the plane perpendicular to the axis of rotation leading from the axis of rotation to the point where the force was applied. r in the moment of inertia I was similarly the *distance* from the axis of rotation of the particular mass m or mass chunk dm. We considered this to be *one dimensional rotation* because the axis of rotation did not change, all rotation was about that one fixed axis.

This is, alas, not terribly general. We started to see that at the end when we talked about the parallel and perpendicular axis theorem and the possibility of *several* "moments of inertia" for a single rigid object around different rotation axes. However, it is really even worse than that. Torque (as we shall see) is a *vector* quantity, and it acts to change *another* vector quantity, the ***angular momentum*** of not just a rigid object, but an arbitrary collection of particles, much as force did when we considered the center of mass. This will have a profound effect on our understanding of certain kinds of phenomena. Let's get started.

We have already identified the axis of rotation as being a suitable "direction" for a one-dimensional torque, and have adopted the right-hand-rule as a means of selecting which of the two directions along the axis will be considered "positive" by convention. We therefore begin by simply generalizing this rule to three dimensions and writing:

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{r}} \times \vec{\boldsymbol{F}} \tag{6.3}$$

where (recall) $\vec{\boldsymbol{r}} \times \vec{\boldsymbol{F}}$ is the ***cross product*** of the two vectors and where $\vec{\boldsymbol{r}}$ is the vector from the ***origin of coordinates or pivot point, not the axis of rotation*** to the point where the force $\vec{\boldsymbol{F}}$ is being applied.

Time for some vector magic! Let's write $\vec{\boldsymbol{F}} = d\vec{\boldsymbol{p}}/dt$ or[136]:

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{r}} \times \vec{\boldsymbol{F}} = \vec{\boldsymbol{r}} \times \frac{d\vec{\boldsymbol{p}}}{dt} = \frac{d}{dt}\left(\vec{\boldsymbol{r}} \times m\vec{\boldsymbol{v}}\right) - \frac{d\vec{\boldsymbol{r}}}{dt} \times m\vec{\boldsymbol{v}} = \frac{d}{dt}\vec{\boldsymbol{r}} \times \vec{\boldsymbol{p}} \tag{6.4}$$

The last term vanishes because $\vec{\boldsymbol{v}} = d\vec{\boldsymbol{r}}/dt$ and $\vec{\boldsymbol{v}} \times m\vec{\boldsymbol{v}} = 0$ for any value of the mass m.

Recalling that $\vec{\boldsymbol{F}} = d\vec{\boldsymbol{p}}/dt$ is Newton's Second Law for vector *translations*, let us *define*:

$$\vec{\boldsymbol{L}} = \vec{\boldsymbol{r}} \times \vec{\boldsymbol{p}} \tag{6.5}$$

as the ***angular momentum vector*** of a particle of mass m and momentum $\vec{\boldsymbol{p}}$ located at a vector position $\vec{\boldsymbol{r}}$ with respect to the origin of coordinates.

[136] I'm using $\frac{d}{dt}\vec{\boldsymbol{r}} \times \vec{\boldsymbol{p}} = \frac{d\vec{\boldsymbol{r}}}{dt} \times \vec{\boldsymbol{p}} + \vec{\boldsymbol{r}} \times \frac{d\vec{\boldsymbol{p}}}{dt}$ to get this, and subtracting the first term over to the other side.

In that case Newton's Second Law for a point mass being *rotated* by a vector torque is:

$$\vec{\boldsymbol{\tau}} = \frac{d\vec{\boldsymbol{L}}}{dt} \tag{6.6}$$

which precisely resembles Newtson's Second Law for a point mass being *translated* by a vector torque.

This is good for a single particle, but what if there are *many* particles? In that case we have to recapitulate our work at the beginning of the center of mass chapter/week.

6.2: Review of the Cross or Vector Product

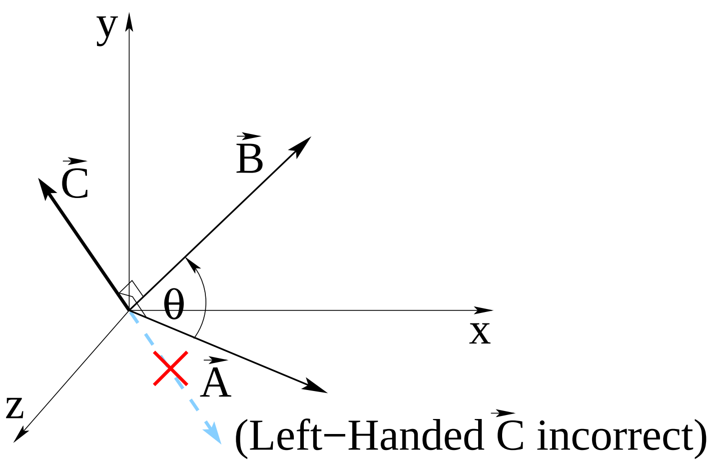


Figure 6.1: The ***cross product*** of two vectors in a three-dimensional (cartesian) space, is at ***right angles*** to the plane in which the two non-colinear vectors lie.

Of the three ways of multiplying a vector with something else covered in the course[137] , the cross product[138] $\vec{C} = \vec{A} \times \vec{B}$ (where all three are vectors[139], also referred to as the vector product, is by far the most difficult for students to have mastered when entering this course. This section is intended to do more than provide students with a broad overview and review. Its real objective is to accomplish this and make remembering and using the cross-product *easy*.

As before, let's start with the standard cartesian representation of the two vectors $\vec{A}$ and vB:

$$\vec{\boldsymbol{A}} = A_x\hat{\boldsymbol{x}} + A_y\hat{\boldsymbol{y}} + A_z\hat{\boldsymbol{z}} \tag{6.7}$$
$$\vec{\boldsymbol{B}} = B_x\hat{\boldsymbol{x}} + B_y\hat{\boldsymbol{y}} + B_z\hat{\boldsymbol{z}} \tag{6.8}$$
$$\tag{6.9}$$

[137]Wikipedia: http://www.wikipedia.org/wiki/Vector Multiplication. There are more... and physics majors might want to think about following links to outer products, geometric algebras, and more. But you won't *need* this right now to do well in this course!

[138]Wikipedia: http://www.wikipedia.org/wiki/Cross Product. The figure in this link shows that the "left-handed" vector not-$\vec{C}$ in 6.1 actually $-\vec{C} = \vec{B} \times \vec{A}$, the negative of $\vec{C} = \vec{A} \times \vec{B}$ as given by the right hand rule.

[139]Sort of... technically $\vec{C}$ is known as a *pseudovector* or *axial vector*, but the difference won't be important in this course.

where $\hat{\boldsymbol{x}}$ is a unit vector in the x direction etc. In figure 6.1, the same $\vec{\boldsymbol{A}}$ and $\vec{\boldsymbol{B}}$ in "arbitrary" directions in a three-dimensional space are illustrated with an angle θ between them, but now I've drawn in the vector:

$$\vec{\boldsymbol{C}} = \vec{\boldsymbol{A}} \times \vec{\boldsymbol{B}} = C_x \hat{\boldsymbol{x}} + C_y \hat{\boldsymbol{y}} + C_z \hat{\boldsymbol{z}}$$

We need to be able to find C_x, C_y, and C_z, given $A_x, A_y, A_z, B_x, B_y, B_z$.

There are several things to note about the vector $\vec{\boldsymbol{C}}$. First, since we want it to represent the same *magnitude* of torque we used in 1D rotation in the previous chapter, its magnitude had better still be:

$$C = AB\sin\theta$$

To indicate a 1D rotation, we need to identify the *axis of rotation*, and do so in a way that preserves the right-hand rule we adopted before by making the axis of rotation into a *vector itself* in 3 dimensions. If our ***right-handed thumb*** aligns with this axis in the direction of $\vec{\boldsymbol{C}}$, then, we know our fingers should naturally curve ***from*** $\boldsymbol{A}$ ***to*** $\boldsymbol{B}$ through the angle $\theta < \pi$! We require a convention (I remind you) so that a space alien on the planet Mongo can be sent coordinates for $\vec{\boldsymbol{A}}$ and $\vec{\boldsymbol{B}}$ (and a specification that right-handed coordinates and products should be used by convention) and they should be able to ship back a right-handed screw that actually goes *in* when twisted to the right/clockwise in a right-handed nut. Without a convention there would be error and chaos[140], and physics has to be *consistent*. This is all illustrated in figure 6.2 *when the two vectors lie in a plane as before*.

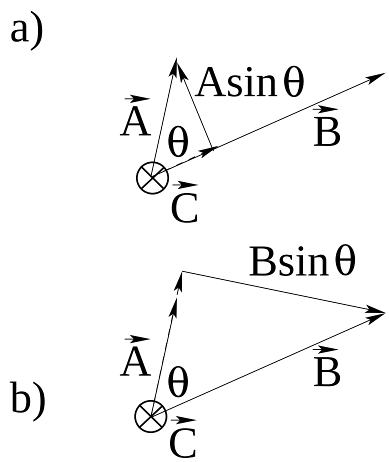


Figure 6.2: Two vectors in a plane have a cross-product perpendicular to the plane in the direction given by the right hand rule.

Note that sometimes – often, even, for simplicity – you can still use $AB\sin\theta$ plus the right-hand rule in this chapter, but you need to be *able* to handle it when you are *indeed* given two arbitrary cartesian vectors and need to get the result in cartesian coordinates. In the general case, however, $\vec{\boldsymbol{A}}$ and $\vec{\boldsymbol{B}}$ are not conveniently in (e.g.) the x-y plane! We don't just get "in" or "out" or "up" or "down" in some already provided coordinate frame! We may not even *know* θ, the angle in between[141] Wowsers! Help!

Relax! No worries, mate! We'll do it step by step, and end up with something as easy as counting 123.

[140]..and not Chaos of the fun sort...

[141]In fact, this is one way to easily *find* the angle θ between two arbitrary vectors.

Literally.

6.2.1: Right Handed Coordinates

Let's start learning about this idea of "handedness" by learning *my* favorite set of rules for determining the direction of a right handed cross product of two vectors that lie in a *convenient* plane (not the other kind yet, but we're getting there):

- First, make sure you are using the ***right right hand*** and not the ***wrong*** right hand! Label your hands if you need to (joke, but no kidding – this is a common mistake when distracted on an exam)!
- Second, ***line the fingers of the right hand*** up with the ***first vector*** (order ***matters***).
- ***Rotate your wrist*** as necessary to ensure that your fingers can ***easily, comfortable, painlessly*** curl into the direction of the ***second vector*** through the **small angle $\boldsymbol{\theta < \pi/2}$** – that is, the orientation that does *not* cause your wrist to break as you try to twist your fingers past π!
- Your ***right-handed thumb*** now indicates the direction **perpendicular to the plane containing the two vectors** that the resultant ***right handed*** cross-product vector points!

From this we immediately conclude one *really important* fact: The cross-product ***changes sign*** when we commute the order of the product. That is:

$$\vec{A} \times \vec{B} = \vec{C} \quad \Rightarrow \quad \vec{B} \times \vec{A} = -\vec{C} = -\vec{A} \times \vec{B} \tag{6.10}$$

Go ahead, try this with the figure 6.2 above.

The cross product is therefore *not commutative* the way ordinary multiplication or the dot product are – order *matters*. We say that the cross-product ***anti***commutes, because reversing order leads to a vector with the same magnitude but the *opposite* direction.

This lets us arrive at the following nifty conclusion:

$$\vec{A} \times \vec{A} = -\vec{A} \times \vec{A} = 0 \tag{6.11}$$

(did you see me swap the order of the vectors there?) because the only vector equal to minus itself is the "zero", or null vector. Since we can multiply any vector by a scalar to change its length, if we let $\vec{B} = \lambda\vec{A}$ for some simply scalar number λ and multiply this equation on both sides by λ, we get:

$$\vec{A} \times \lambda\vec{A} = \vec{A} \times \vec{B} = \lambda \times 0 = 0 \tag{6.12}$$

or (in English): "Any two parallel or antiparallel vectors have a cross product of zero".

Similarly if we have two *perpendicular* vectors $\vec{A}$ and $\vec{B}$ (so the angle between them is $\pm\pi/2$):

$$|\vec{A} \times \vec{B}| = |AB\sin(\pm\pi/2)| = |\pm AB| = AB \tag{6.13}$$

and if the vectors have length 1 (unit vectors), this implies (again, in English): "The cross product of two perpendicular unit vectors is therefore *also* a unit vector" (with direction given by the right hand rule).

Now, you may not have realized it, but there are actually ***two distinct ways*** to draw a 3D cartesian coordinate frame. They are illustrated in figure 6.3.

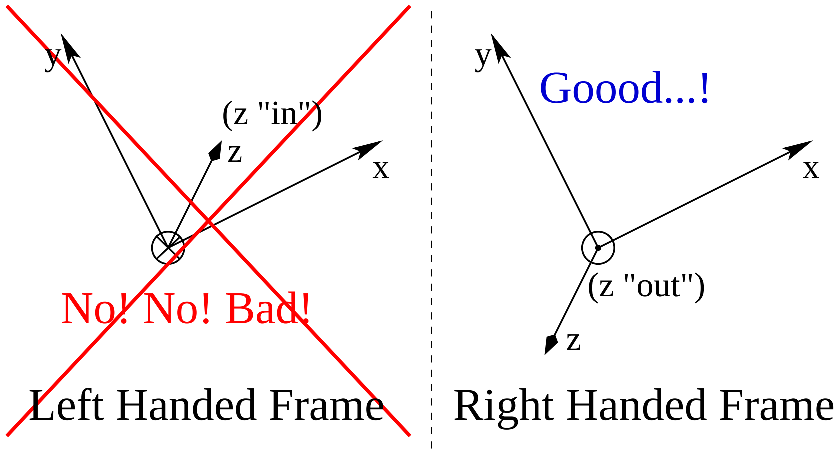


Figure 6.3: The frame on the left is ***left-handed*** (appropriately). The frame on the right is ***right-handed***. Can you see the difference? You ***cannot rotate one so that it matches the other!*** They are *completely distinct!*

It is *easy* to see why the two frames are fundamentally different. In the first frame on the left, if you line your *left hand fingers* up with $\hat{x}$ and curl them through $\pi/2$ to line up with $\hat{\boldsymbol{y}}$, your ***left hand thumb*** points in the $\hat{z}$ direction. These coordinates are then said to be left handed. As the bright $\times$ and matching text clearly indicate, this is a coordinate frame that we must never, ever use in this course – by convention!

In the frame on the right, if you line your *right hand fingers* up with $\hat{x}$ and curl them through $\pi/2$ to line up with $\hat{\boldsymbol{y}}$, your ***right hand thumb*** points in the $\hat{z}$ direction. These coordinates are then said to be left handed. As the cool blue Goood indicates, we must now form a pact to ***only use right handed coordinates*** as illustrated in the ***right handed frame*** from now on!

This, then, we all collectively swear, on penalty of, well, getting things *wrong* on quizzes and exams or getting back screws manufactured overseas that are lefty-tighty and righty-loosey, which doen't even sort-of-rhyme and is likely to get you fired. Or (if you prefer biochemistry) giving somebody a *right-handed* drug in place of the left-handed version that is actually compatible with human life or vice versa[142] which might end up with them being dead – or worse[143].

As you can see by direct inspection, if unit vectors $\hat{x}$ and $\hat{y}$ are perpendicular to each other, in a right handed coordinate frame

$$|\hat{\boldsymbol{z}}| = |\hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}}| = 1 \times 1 \times \sin \pi/2 = 1$$

and ***in that order***:

$$\hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}} = \hat{\boldsymbol{z}} \tag{6.14}$$

[142] Wikipedia: http://www.wikipedia.org/wiki/Chirality (chemistry). OK, so I admit it, in biology relevant/safe organic molecules are often *left* handed so you should probably use your left hand there. The point is that you should be able to see and understand the difference, because molecular isomers of the opposite chirality are not, actually, interchangeable in function!

[143] https://www.scientificamerican.com/article/mirror-molecules-nature-not-ambidextrous/ Except that it is really even more complicated than that. This is an intersection of physics, chemistry and biology that is most interesting.

In fact, we use precisely this relation to ***define*** a right handed frame. It is one where $\hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}} = \hat{\boldsymbol{z}}$ in that precise order! A left handed frame has $\hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}} = -\hat{\boldsymbol{z}}$ (check it in the figure above, using your *right* hand only from now on).

Use your right hand and the (right hand) figure above to determine what $\hat{\boldsymbol{x}} \times \hat{\boldsymbol{z}}$ is. $-\hat{\boldsymbol{y}}$, right? Can you now write down *all* of the possible products of unit vectors (there are 9 of them)?

Well from the parallel rules above:

$$|\hat{\boldsymbol{x}} \times \hat{\boldsymbol{x}}| = |\hat{\boldsymbol{y}} \times \hat{\boldsymbol{y}}| = |\hat{\boldsymbol{y}} \times \hat{\boldsymbol{z}}| = 0$$

and each possible combination is easy to get using the right hand rules give above. In the end, you should get the following table (in array format):

$$\begin{array}{lll} \hat{\boldsymbol{x}} \times \hat{\boldsymbol{x}} = 0 & \hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}} = \hat{\boldsymbol{z}} & \hat{\boldsymbol{x}} \times \hat{\boldsymbol{z}} = -\hat{\boldsymbol{y}} \\ \hat{\boldsymbol{y}} \times \hat{\boldsymbol{x}} = -\hat{\boldsymbol{z}} & \hat{\boldsymbol{y}} \times \hat{\boldsymbol{y}} = 0 & \hat{\boldsymbol{y}} \times \hat{\boldsymbol{z}} = \hat{\boldsymbol{x}} \\ \hat{\boldsymbol{z}} \times \hat{\boldsymbol{x}} = -\hat{\boldsymbol{y}} & \hat{\boldsymbol{z}} \times \hat{\boldsymbol{y}} = -\hat{\boldsymbol{x}} & \hat{\boldsymbol{z}} \times \hat{\boldsymbol{z}} = 0 \end{array}$$

This looks really hard to remember, but actually, it is *easy*.

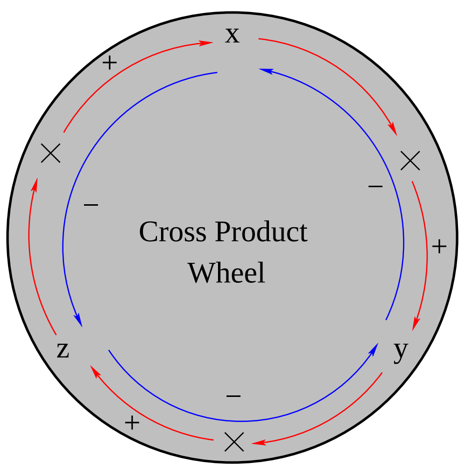


Figure 6.4: The "Wheel of Fortune" – you are *fortunate* to be able to make the cross product as easy as ABC or 123 or – xyz...

The *entire multiplication table* can be reduced to remembering the "wheel of fortune" pictured in 6.4. Note that I drew draw $\hat{\boldsymbol{x}}$ at the "twelve o'clock position" on the wheel, $\hat{\boldsymbol{y}}$ at the four o'clock position, and $\hat{\boldsymbol{z}}$ at the eight o'clock position. Starting at $\hat{\boldsymbol{x}}$ and going around the wheel in the *clockwise* direction – called "cyclic order" – we get $\hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}} = +\hat{\boldsymbol{z}}$. This is our definition of a right handed frame, all good. But look! If you start at $\hat{\boldsymbol{y}}$ and go clockwise, the wheel tells you that $\hat{\boldsymbol{y}} \times hz = +\hat{\boldsymbol{x}}$! If you start at $\hat{\boldsymbol{z}}$ and go clockwise, the wheel tells you that $\hat{\boldsymbol{z}} \times \hat{\boldsymbol{x}} = +\hat{\boldsymbol{y}}$. And if you look at the multiplication table above, ***this is exactly correct!***

Because the product is anticommutative, reversing direction just makes each product negative. This means that we can use the same wheel but go around it *anticlockwise* (counterclockwise) and put in a minus sign on each result. Starting with $\hat{\boldsymbol{x}}$ and following the blue arrows around, we get $\hat{\boldsymbol{x}} \times \hat{\boldsymbol{z}} = -\hat{\boldsymbol{y}}$. This is also exactly correct! You should now be able to use the wheel counterclockwise to verify the remaining two nonzero products: $\hat{\boldsymbol{z}} \times \hat{\boldsymbol{y}} = -\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}} \times \hat{\boldsymbol{x}} = -\hat{\boldsymbol{z}}$!

We add the trivial rule that any unit vector cross with itself is zero, and we have all nine products of unit vectors and are ready to assemble a full cartesian cross-product just about as fast as we can right it down. We only need two more of the rules of multiplication – the cross product is ***associative*** and *distributive*, as long as you *preserve the order of products* as you factor or multiply – vector components on the left have to *stay* on the left and vice versa as you multiply things out!

Let's do it! In full, gory detail:

$$\vec{\boldsymbol{A}} \times \vec{\boldsymbol{B}} = \quad A_x B_x \cancel{\hat{\boldsymbol{x}} \times \hat{\boldsymbol{x}}} + A_x B_y (\hat{\boldsymbol{x}} \times \hat{\boldsymbol{y}}) + A_x B_z (\hat{\boldsymbol{x}} \times \hat{\boldsymbol{z}}) \tag{6.15}$$

$$+ A_y B_x (\hat{\boldsymbol{y}} \times \hat{\boldsymbol{x}}) + A_y B_y \cancel{\hat{\boldsymbol{y}} \times \hat{\boldsymbol{y}}} + A_y B_z (\hat{\boldsymbol{y}} \times \hat{\boldsymbol{z}}) \tag{6.16}$$

$$+ A_z B_x (\hat{\boldsymbol{z}} \times \hat{\boldsymbol{x}}) + A_z B_y (\hat{\boldsymbol{z}} \times \hat{\boldsymbol{y}}) + A_z B_z \cancel{\hat{\boldsymbol{z}} \times \hat{\boldsymbol{z}}} \tag{6.17}$$

All of the non-diagonal dot products of unit vectors are zero (as illustrated in the cancellations above)! We can fill in the products in parentheses using the wheel – no peeking, now, just remember to use the cyclic permutations of xyz in clockwise order with a positive sign, and reverse the sign if you do them in the counterclockwise order. You should get:

$$\vec{\boldsymbol{A}} \times \vec{\boldsymbol{B}} = A_x B_y (\hat{\boldsymbol{z}}) + A_x B_z (-\hat{\boldsymbol{y}}) \tag{6.18}$$

$$+ A_y B_x (-\hat{\boldsymbol{z}}) + A_y B_z (\hat{\boldsymbol{x}}) \tag{6.19}$$

$$+ A_z B_x (\hat{\boldsymbol{y}}) + A_z B_y (-\hat{\boldsymbol{x}}) \tag{6.20}$$

I deliberately left this unsorted, but we can clearly collect the terms with the same unit vectors to get (in order):

$$\vec{\boldsymbol{A}} \times \vec{\boldsymbol{B}} = (A_y B_z - A_z B_y)\hat{\boldsymbol{x}} + (A_z B_x - A_x B_z)\hat{\boldsymbol{y}} + (A_x B_y - A_y B_x)\hat{\boldsymbol{z}} \tag{6.21}$$

only *that's now how I did it!* Too hard, too hard!

What I *actually* did was mentally sort $xyz- > yzx- > zxy$ (the three cyclic permutations of xyz in order). I then thought "I need the $\hat{\boldsymbol{x}}$ term first, that comes from y times z in that order" (second in the list). I then wrote down while saying in my mind, in English: "Open parentheses A sub $\boldsymbol{y}$ times B sub $\boldsymbol{z}$, then put in a minus sign and reverse the indices, - A sub $\boldsymbol{z}$ times B sub $\boldsymbol{y}$, then close parentheses, times $\hat{\boldsymbol{x}}$" to write the first term. To get the $\hat{\boldsymbol{y}}$ term I said to myself in Mathematese (that is, with the same English, from the wheel zx to y) "$(A_z B_x - A_x B_z)\hat{\boldsymbol{y}}$".

See if you can do the $\hat{\boldsymbol{z}}$ term on your own, without looking at anything. It is the easiest one – it follows from xy to z!

I actually did this without looking at *anything*, certainly not by multiplying it out and factoring it! Once you get the hang of this, you can write down the cartesian form of the cross product without even blinking, or using some bizarre form of a determinant, or panicking. It is literally as easy as 1-2-3 (or at least, as $123- > 231- > 312$ in cyclic order). These rules can be *algebraically* represented as sums over the *Levi-Civita fully antisymmetric third rank unit tensor* times the indexed components of the vectors, but that is beyond the scope of *this* course. You can learn more about it if you are a math or physics major by following links in the Cross Product wikilink above.

There are only a couple more points to make about the cross product that I include as entrés into more advanced topics. They are *not* necessary for this course, but are pretty cool. The first is that you can use it (or the dot product, likely easier) to *find* the angle in between two arbitrary vectors. This follows from:

$$\cos\theta = \frac{\vec{\boldsymbol{A}}\cdot\vec{\boldsymbol{B}}}{AB} \qquad \sin\theta = \frac{|\vec{\boldsymbol{A}}\times\vec{\boldsymbol{B}}|}{AB}$$

Using these, one can invent something called *Gram-Schmidt orthogonalization*[144] that is a clever way of taking a set of arbitrary vectors that *span* a vector space but may not have unit length and turn them into a orthogonal basis – a set of unit vectors that span the space and can be scaled to represent any vector. This in turn is a major *practical* component of Gauss elimination in *linear algebra* – a required subject for many students of math, physics, computer science, statistics and certain types of engineering and de facto the way that I usually suggest that one solve simple linear algebra problems in this course.

The last one is that if $\vec{\boldsymbol{A}}$ and $\vec{\boldsymbol{B}}$ both have units of length and are not parallel, the magnitude of the cross product $|\vec{\boldsymbol{A}}\times\vec{\boldsymbol{B}}|$ is the ***area of the parallelogram bounded by $\vec{A}$ and B!*** It is therefore sometimes called the *areal* cross product. If you have a third length vector $\vec{\boldsymbol{C}}$, the volume of the parallelopiped bounded by all three is $|\vec{\boldsymbol{C}}\cdot(\vec{\boldsymbol{A}}\times\vec{\boldsymbol{B}})|$. One can think *geometrically* of the cross product, then, as a kind of *directed area*. One can make the actual *direction* of this area very abstract and algebraic, so it works independent (more or less) of the dimensionality of the space! This observation is one of things that is used to derive a topic called *geometric algebra*[145] that physics and math majors, at least, might find very interesting indeed!

6.3: Total Torque

In figure 6.5 a small collection of (three) particles is shown, each with both "external" forces $\vec{\boldsymbol{F}}_i$ and "internal" forces $\vec{\boldsymbol{F}}_{ij}$ portrayed. The forces and particles do *not* necessarily live in a plane – we simply cannot see their z-components. Also, this picture has just enough particles illustrated to help us visualize, but be thinking of adding more particles with indices $i = ..., 4, 5, 6...N$ (for any N) as we proceed.

Let us write $\vec{\boldsymbol{\tau}} = d\vec{\boldsymbol{L}}/dt$ for each particle and sum the whole thing up, much as we did for $\vec{\boldsymbol{F}} = d\vec{\boldsymbol{p}}/dt$ in chapter/week 4:

$$\begin{aligned}\vec{\boldsymbol{\tau}}_{\rm tot} = \sum_i \vec{\boldsymbol{r}}_i \times \left(\vec{\boldsymbol{F}}_i + \sum_{j\neq i}\vec{\boldsymbol{F}}_{ij}\right) &= \frac{d}{dt}\sum_i \vec{\boldsymbol{r}}_i\times\vec{\boldsymbol{p}}_i = \frac{d\vec{\boldsymbol{L}}_{\rm tot}}{dt} \\ \left(\sum_i \vec{\boldsymbol{r}}_i\times\vec{\boldsymbol{F}}_i\right) + \left(\sum_i\sum_{j\neq i}\vec{\boldsymbol{r}}_i\times\vec{\boldsymbol{F}}_{ij}\right) &= \frac{d}{dt}\sum_i \vec{\boldsymbol{r}}_i\times\vec{\boldsymbol{p}}_i \end{aligned} \tag{6.22}$$

Consider the term that sums the *internal* torques, the torques produced by the internal forces between the particles, for a particular pair (say, particles 1 and 2) and use good old N3,

[144] Wikipedia: http://www.wikipedia.org/wiki/Gram-Schmidt process.

[145] Wikipedia: http://www.wikipedia.org/wiki/Geometric algebra. A way, way cool subject and *very* tightly coupled to physics from introductory mechanics and E&M all the way through quantum field theory.

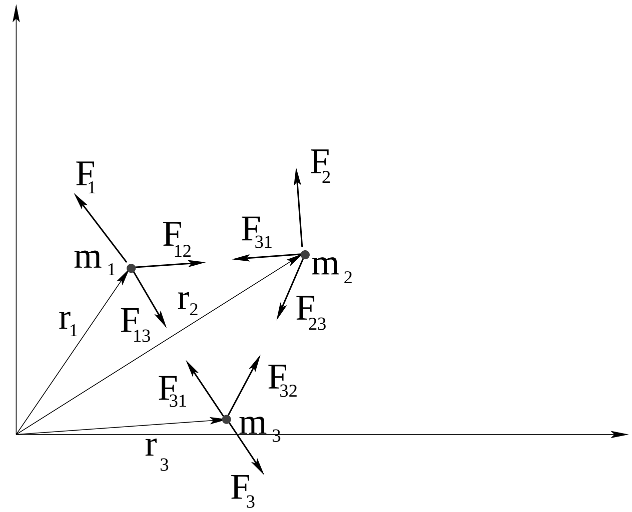


Figure 6.5: The coordinates of a small collection of particles, just enough to illustrate how internal torques work out.

$\vec{\boldsymbol{F}}_{21} = -\vec{\boldsymbol{F}}_{12}$:

$$\vec{\boldsymbol{r}}_1 \times \vec{\boldsymbol{F}}_{12} + \vec{\boldsymbol{r}}_2 \times \vec{\boldsymbol{F}}_{21} = \vec{\boldsymbol{r}}_1 \times \vec{\boldsymbol{F}}_{12} - \vec{\boldsymbol{r}}_2 \times \vec{\boldsymbol{F}}_{12} = (\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2) \times \vec{\boldsymbol{F}}_{12} = 0 \quad (!) \tag{6.23}$$

because $\vec{\boldsymbol{r}}_{12} = \vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2$ is ***parallel or antiparallel*** to $\vec{\boldsymbol{F}}_{12}$ and the cross product of two vectors that are parallel or antiparallel is *zero*.

Obviously, the same algebra holds for *any* internal force pair so that:

$$\left(\sum_i \sum_{j \neq i} \vec{\boldsymbol{r}}_i \times \vec{\boldsymbol{F}}_{ij} \right) = 0 \tag{6.24}$$

and

$$\vec{\boldsymbol{\tau}}_{\rm tot} = \sum_i \vec{\boldsymbol{r}}_i \times \vec{\boldsymbol{F}}_i = \frac{d}{dt} \sum_i \vec{\boldsymbol{r}}_i \times \vec{\boldsymbol{p}}_i = \frac{d\vec{\boldsymbol{L}}_{\rm tot}}{dt} \tag{6.25}$$

where $\vec{\boldsymbol{\tau}}_{\rm tot}$ is the sum of only the *external* torques – the *internal* torques cancel.

The physical meaning of this cancellation of internal torques is simple – just as you cannot lift yourself up by your own bootstraps, because internal opposing forces acting along the lines connecting particles can never alter the velocity of the center of mass or the total momentum of the system, you cannot exert a torque on yourself and alter your own total angular momentum – only the total *external* torque acting on a system can alter its total angular momentum.

Wait, what's that? An isolated system (one with no net force or torque acting) must have a *constant angular momentum?* Sounds like a conservation law to me...

6.3.1: The Law of Conservation of Angular Momentum

We've basically done everything but write this down above, so let's state it clearly in both words and algebraic notation. First in words:

> ***If and only if*** the total vector torque acting on any system of particles is zero, ***then*** the total angular momentum of the system is a constant vector.

In equations it is even more succinct:

$$\text{If and only if } \vec{\tau}_{\rm tot} = 0 \text{ then } \vec{L}_{\rm tot} = \vec{L}_{\rm initial} = \vec{L}_{\rm final} = \text{ a constant vector} \tag{6.26}$$

Note that (like the Law of Conservation of Momentum) this is a *conditional* law – angular momentum is conserved *if and only if* the net torque acting on a system is zero (so if angular amomentum *is* conserved, you may conclude that the total torque is zero as that is the only way it could come about).

Just as was the case for Conservation of Momentum, our primary use at this point for Conservation of Angular Momentum will be to help analyze ***collisions***. Clearly the internal forces in two-body collisions in the ***impulse approximation*** (which allows us to ignore the *torques* exerted by external forces during the tiny time Δt of the impact) can exert no net torque, therefore we expect ***both*** linear momentum ***and*** angular momentum to be conserved during a collision.

Before we proceed to analyze collisions, however, we need to understand angular momentum (the conserved quantity) in more detail, because it, like momentum, is a *very important quantity* in nature. In part this is because many elementary particles (such as quarks, electrons, heavy vector bosons) and many microscopic composite particles (such as protons and neutrons, atomic nuclei, atoms, and even molecules) can have a net *intrinsic* angular momentum, called *spin*[146] .

This spin angular momentum is not *classical* and does not arise from the physical motion of mass in some kind of path around an axis – and hence is largely beyond the scope of this class, but we certainly need to know how to evaluate and alter (via a torque) the angular momentum of macroscopic objects and collections of particles as they rotate about fixed axes.

6.4: The Angular Momentum of a Symmetric Rotating Rigid Object

One very important aspect of both vector torque and vector angular momentum is that $\vec{r}$ in the definition of both is ***measured from a pivot that is a single point***, not measured from a pivot *axis* as we imagined it to be last week when considering only one dimensional rotations. We would very much like to see how the two general descriptions of rotation are related, though, especially as at this point we should *intuitively* feel (given the strong correspondance between one-dimensional linear motion equations and one-dimensional angular motion equations) that something like $L_z = I\Omega_z$ ought to hold to relate angular momentum to the moment of inertia. Our intuition is mostly correct, as it turns out, but things are a little more complicated than that.

From the derivation and definitions above, we expect angular momentum $\vec{L}$ to have three components just like a spatial vector. We also expect $\vec{\Omega}$ to be a vector (that points in the direction of the right-handed axis of rotation that passes through the pivot point). We expect there

[146]Wikipedia: http://www.wikipedia.org/wiki/Spin (physics). Physics majors should probably take a peek at this link, as well as chem majors who plan to or are taking physical chemistry. I foresee the learning of Quantum Theory in Your Futures, and believe me, you *want* to preload your neocortex with lots of quantum cartoons and glances at the algebra of angular momentum in quantum theory ahead of time...

to be a linear relationship between angular velocity and angular momentum. Finally, based on our observation of an extremely consistent analogy between quantities in one dimensional linear motion and one dimensional rotation, we expect the moment of inertia to be a quantity that transforms the angular velocity into the angular momentum by some sort of multiplication.

To work out all of these relationships, we need to start by *indexing* the particular axes in the coordinate system we are considering with e.g. $a = x, y, z$ and label things like the components of $\vec{\boldsymbol{L}}$, $\vec{\boldsymbol{\Omega}}$ and I with a. Then $L_{a=z}$ is the z-component of $\vec{\boldsymbol{L}}$, $\Omega_{a=x}$ is the x-component of $\vec{\boldsymbol{\Omega}}$ and so on. This is simple enough.

It is not so simple, however, to generalize the moment of inertia to three dimensions. Our simple one-dimensional *scalar* moment of inertia from the last chapter clearly depends on the particular axis of rotation chosen! For rotations around the (say) z-axis we needed to sum up $I = \sum_i m_i r_i^2$ (for example) where $r_i = \sqrt{x_i^2 + y_i^2}$, and these components were clearly all *different* for a rotation around the x-axis or a z axis through a different pivot (perpendicular or parallel axis theorems). These were still the easy cases – as we'll see below, things get really complicated when we rotate even a symmetric object around an axis that is not an axis of symmetry of the object!

Indeed, what we have been evaluating thus far is more correctly called the *scalar* moment of inertia, the moment of inertia evaluated around a particular "obvious" one-dimensional axis of rotation where one or both of two ***symmetry conditions*** given below are satisfied. The moment of inertia of a general object in some coordinate system is more generally described by the *moment of inertia tensor* I_{ab}. Treating the moment of inertia tensor correctly is beyond the scope of this course, but math, physics or engineering students are well advised to take a peek at the Wikipedia article on the moment of inertia[147] to at least get a glimpse of the mathematically more elegant and correct version of what we are covering here.

Here are the two conditions and the result. Consider a ***particular pivot point at the origin of coordinates*** and ***right handed rotation around an axis in the*** a***th direction*** of a coordinate frame with this origin. Let the ***plane of rotation*** be the plane perpendicular to this axis that contains the pivot/origin. There isn't anything particularly mysterious about this – think of the $a = z$-axis being the axis of rotation, with positive in the right-handed direction of $\vec{\boldsymbol{\Omega}}$, and with the x-y plane being the plane of rotation.

In this coordinate frame, ***if*** the mass distribution has:

- ***Mirror symmetry across the axis of rotation*** and/or
- ***Mirror symmetry across the plane of rotation***,

we can write:

$$L = L_a = I_{aa}\Omega_a = I\Omega \tag{6.27}$$

where $\vec{\boldsymbol{\Omega}} = \Omega_a \hat{\boldsymbol{a}}$ points in the (right handed) direction of the axis of rotation and where:

$$I = I_{aa} = \sum_i m_i r_i^2 \quad \text{or} \quad \int r^2 dm \tag{6.28}$$

[147]Wikipedia: http://www.wikipedia.org/wiki/Moment_of_inertia#Moment_of_inertia_tensor. This is a link to the middle of the article and the tensor part, but even introductory students may find it useful to review the *beginning* of this article.

with r_i or r the distance from the a-axis of rotation as usual.

Note that "mirror symmetry" just means that if there is a chunk of mass or point mass in the rigid object on one side of the axis or plane of rotation, there is an equal chunk of mass or point mass in the "mirror position" on the exact opposite of the line or plane, for every bit of mass that makes up the object. This will be illustrated in the next section below, along with *why* these rules are needed.

In other words, the scalar moments of inertia I we evaluated last chapter are just the diagonal parts of the moment of inertia tensor $I = I_{aa}$ for the coordinate direction a corresponding to the axis of rotation. Since we aren't going to do much – well, we aren't going to do *anything* – with the non-diagonal parts of I in this course, from now on I will just write the scalar moment I where I really mean I_{aa} for some axis a such that at least one of the two conditions above are satisfied, but math/physics/engineering students, at least, should try to remember that it really ain't so[148].

All of the (scalar) I's we computed in the last chapter satisfied these symmetry conditions:

- A ring rotating about an axis through the center perpendicular to the plane of the ring has both symmetries. So does a disk.
- A rod rotating about one end in the plane perpendicular to $\vec{\Omega}$ has mirror symmetry in the plane but not mirror symmetry across the axis of rotation.
- A hollow or filled sphere have both symmetries.
- A disk around an axis off to the side (evaluated using the parallel axis theorem) has the plane symmetry.
- A disk around an axis that lies in the plane of the disk with a pivot in the perpendicular plane through the center of the disk has at least the planar symmetry relative to the perpendicular plane of rotation.

and so on. Nearly all of the problems we consider in this course will be sufficiently symmetric that we can use:

$$L = I\Omega \tag{6.29}$$

with the pivot and direction of the rotation and the symmetry of the object with respect to the axis and/or plane of rotation "understood".

6.4.1: Spin and Orbital Angular Momentum

As we have seen, even a point mass moving in a straight line has a *kind* of angular momentum even though it isn't really "rotating" about a pivot that is not along its straight line trajectory, merely sweeping out a total angle of π angle as it moves from $-\infty$ to ∞ past the pivot. Of

[148] Just FYI, in case you care: The correct rule for computing $\vec{L}$ from $\vec{\Omega}$ is

$$L_a = \sum_b I_{ab}\Omega_b$$

for $a, b = x, y, z$.

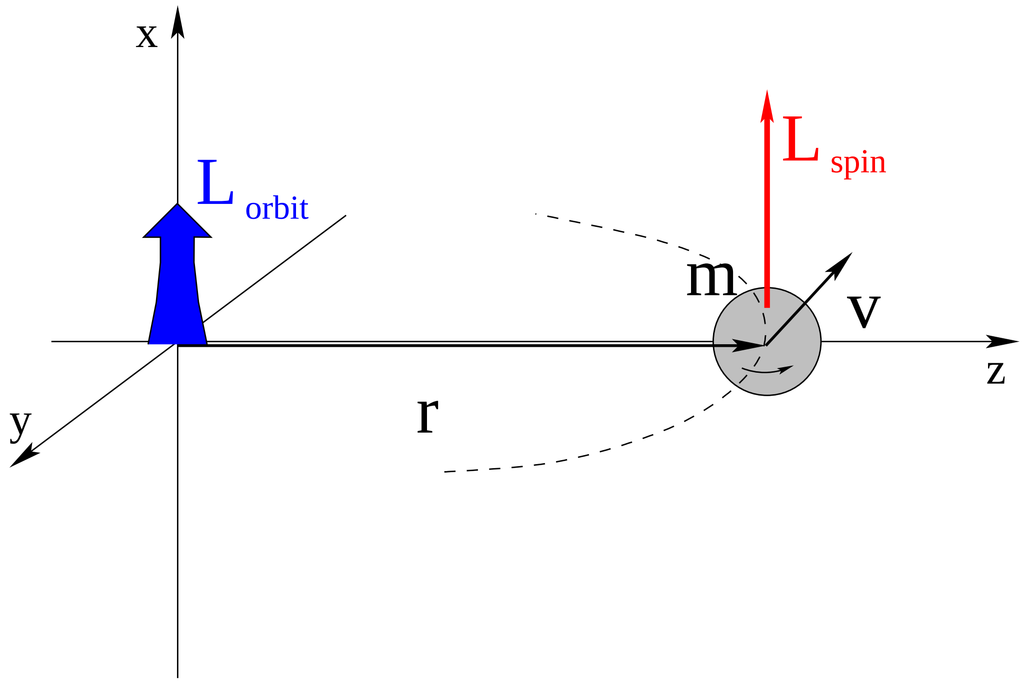


Figure 6.6: The Earth has both ***spin*** angular momentum as it spins daily on its own axis and *orbital* angular momentum as it orbits the Sun – relative to a Sun-centered pivot!

course, many particles – for example, the *planets* – have *actual* angular momentum as they orbit the the Sun (relative to a pivot in the Sun). On the other hand, *rigid bodies* can have angular momentum all by themselves as they rotate uniformly around an axis through their own center of mass. Finally, we have seen that as we choose different pivots for a rolling object, at the very least our rotational *kinetic energy* picks up contributions from *both* the translational angular momentum *of* the center of mass as if it is a particle *and* the rotational angular momentum of the object as it rotates around its own center of mass via the parallel axis theorem.

All of these things together suggest that our description so far is incomplete. The rolling constraint is great for rolling objects, but what about objects like planets orbiting the Sun or electrons orbiting an atomic nucleus? Planets treated like *particles* clearly have angular momentum relative to a pivot located at the Sun, but – don't they also *rotate*?

They do. They just don't necessarily rotate around their *own* axis with the same angular frequency with which they revolve around the central (gravitational) attractor, the Sun. They have angular momentum – even relative to a Sun-centered pivot – *two ways!* Worse, there is nothing that forces the two *kinds* of angular momentum they can have to *point in the same direction*, and in fact they often do not!

Take the case of the Earth, which revolves around the Sun once every $365\frac{1}{4}$ days (and hence has particle-like angular momentum) perpendicular to the "ecliptic" plane of the solar system where most of the planets are to be found). It also rotates on its own axis once every 24 hour, and hence has (approximately) the angular momentum of a solid sphere of mass, and *this* angular momentum is *not at all parallel to the first*, as the Earth's axis is *tipped by 26°* relative to the angular momentum of its orbit.

This sort of situation, pictured above in figure 6.6 above, is so common that we have invented special adjectives to differentiate the two kinds of angular momentum, which are *not* that likely to be related by the parallel axis theorem (although in some cases, like the Moon orbiting the Earth in a "tidally locked" orbit, they may be). We call the part arising from revolution *around* a central attractor ***orbital angular momentum*** and give it its own symbol, $\vec{L}$, and the part that arises from rotation around an axis through its own center of mass ***spin angular momemtum*** and give it the symbol $\vec{S}$. The total angular momentum is then given by:

$$\vec{J} \text{ or } \vec{L}_{\rm tot} = \vec{L} + \vec{S}$$

(the $\vec{J}$ symbol is common in nuclear physics and atomic physics, but as far as I know isn't usually used to describe planetary angular momentum).

Let us take a quick tour, then, of the angular momentum we expect in these cases. A handful of examples should suffice, where I will try to indicate the correct direction as well as show the "understood" scalar result.

Example 6.4.1: Angular Momentum of a Point Mass Moving in a Circle

For a point mass moving in a circle of radius r in the x-y plane, we have the ***planar symmetry***. $\vec{\Omega} = \Omega\hat{z}$ is in the z-direction, and $I = I_{zz} = mr^2$. The angular momentum in this direction is:

$$L = L_z = (\vec{r} \times \vec{p})_z = mvr = mvr \cdot \left(\frac{r}{r}\right) = (mr^2)\frac{v}{r} = I\Omega \qquad (6.30)$$

The ***direction*** of this angular momentum is most easily found by using a variant of the right hand rule. Let the fingers of your right hand curl around the axis of rotation in the direction of the motion of the mass. Then your thumb points out the direction. You should verify that this gives the ***same result*** as using $\vec{L} = \vec{r} \times \vec{p}$, always, but this "grasp the axis" rule is much easier and faster to use, just grab the axis with your fingers curled in the direction of rotation and your thumb has got it.

Example 6.4.2: Angular Momentum of a Rod Swinging in a Circle

To compute the angular momentum of a rod rotating in a plane around a pivot through one end, we choose coordinates such that the rod is in the x-y plane, rotating around z, and has mass M and length l (note that it is now tricky to call its length L as that's also the symbol for angular momentum, sigh). From the previous example, each little "point-like" bit of mass in the rod dm has an angular momentum of:

$$dL_z = |\vec{r} \times d\vec{p}| = r(dm\ v) = \left(r^2 dm\right)\left(\frac{v}{r}\right) = dI\ \Omega \qquad (6.31)$$

so that if we integrate this as usual from 0 to l, we get:

$$|\vec{L}| = L_z = \frac{1}{3}Ml^2\Omega = I\Omega \qquad (6.32)$$

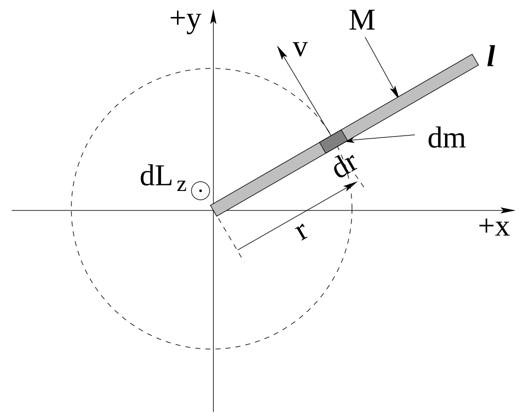


Figure 6.7: The geometry of a rod of mass M and length L, rotating around a pivot through the end in the x-y plane.

Example 6.4.3: Angular Momentum of a Rotating Disk

Suppose a disk is rotating around its center of mass in the x-y plane of the disk. Then using *exactly the same argument as before*:

$$L_z = \int r^2 dm\Omega = I\Omega = \frac{1}{2}MR^2\Omega \tag{6.33}$$

The disk is symmetric, so if we should be rotating it like a spinning coin or poker chip around (say) the x axis, we can also find (using the perpendicular axis theorem to find I_x):

$$L_x = I_x\Omega = \frac{1}{4}MR^2\Omega \tag{6.34}$$

and you *begin* to see why the direction labels are necessary. A disk has a *different scalar moment of inertia* about different axes through the same pivot point. Even when the symmetry is obvious, we may still need to label the result or risk confusing the previous two results!

We're not done! If we attach the disk to a massless string and swing it around the z axis at a distance ℓ from the center of mass, we can use the *parallel* axis theorem and find that:

$$L_z^{\rm new} = I_z^{\rm new}\Omega = (M\ell^2 + \frac{1}{2}MR^2)\Omega \tag{6.35}$$

That's *three* results for a single object, and of course we can apply the parallel axis theorem to the x-rotation or y-rotation as well! The $L_i = I_i\Omega$ (for $i = x, y, z$) result works, but the direction of $\vec{L}$ and $\vec{\Omega}$ as well as the value of the scalar moment of inertia I used will vary from case to case, so you should carefully label even the magnitude of angular momentum whenever there is any chance of confusion just to avoid making mistakes!

6.5: Angular Momentum Conservation

We have derived (trivially) the Law of Conservation of Angular Momentum: When the total external torque acting on a systems is zero, the total angular momentum of the system is constant, that is, conserved. As you can imagine, this is a powerful concept we can use to understand many everyday phenomena and to solve many problems, both very simple conceptual ones and very complex and difficult ones.

The simplest application of this concept comes, now that we understand well the relationship between the scalar moment of inertia and the angular momentum, in systems where the moment of inertia of the system can *change over time* due to strictly *internal* forces. We will look at two particular example problems in this genre, deriving a few very useful results along the way.

Example 6.5.1: The Spinning Professor

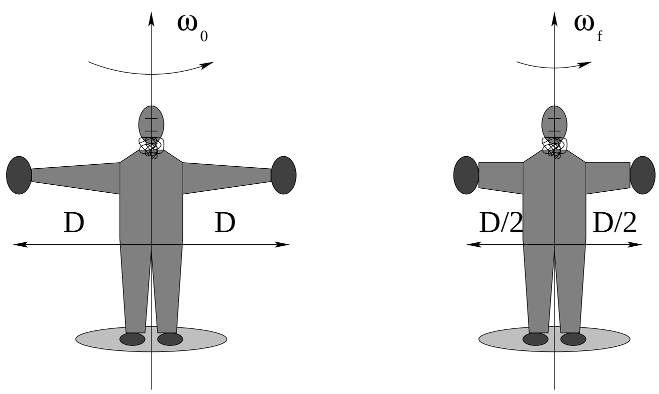


Figure 6.8: A professor stands on a freely pivoted platform at rest (total moment of inertia of professor and platform I_0) with two large masses m held horizontally out at the side a distance D from the axis of rotation, initially rotating with some angular velocity Ω_0.

A professor stands on a freely pivoted platform at rest with large masses held horizontally out at the side. A student gives the professor a push to start the platform and professor and masses rotating around a vertical axis. The professor then pulls the masses in towards the axis of rotation, reducing their contribution to the total moment of inertia as illustrated in figure **??**

If the moment of inertia of the professor and platform is I_0 and the masses m (including the arms' contribution) are held at a distance D from the axis of rotation and the initial angular velocity is Ω_0, what is the final angular velocity of the system Ω_f when the professor has pulled the masses in to a distance $D/2$?

The platform is *freely pivoted* so it exerts no external torque on the system. Pulling in the masses exerts no *external* torque on the system (although it may well exert a torque on the masses themselves as they transfer angular momentum to the professor). ***The angular momentum of the system is thus conserved***.

Initially it is (in this highly idealized description)

$$L_i = I_i \Omega_0 = (2mD^2 + I_0)\Omega_0 \tag{6.36}$$

Finally it is:

$$L_f = I_f \Omega_f = (2m\left(\frac{D}{2}\right)^2 + I_0)\Omega_f = (2mD^2 + I_0)\Omega_0 = L_i \tag{6.37}$$

Solving for Ω_f:

$$\Omega_f = \frac{(2mD^2 + I_0)}{(2m\left(\frac{D}{2}\right)^2 + I_0)}\Omega_0 \tag{6.38}$$

From this all sorts of other things can be asked and answered. For example, what is the initial kinetic energy of the system in terms of the givens? What is the final? How much work did the professor do with his arms?

Note that this is *exactly how* ice skaters speed up their spin when performing their various nifty moves – start spinning with arms and legs spread out, then draw them in to spin up, extend them to slow down again. It is how high-divers control their rotation. It is how neutron stars spin up as their parent stars explode. It is part of the way cats manage to always land on their feet – for a value of the world "always" that really means "usually" or "mostly"[149].

Although there are more general ways of a system of particles altering its own moment of inertia, a fairly common way is indeed through the application of what we might call *radial* forces. Radial forces are a bit special and worth treating in the context of angular momentum conservation in their own right.

6.5.1: Radial Forces and Angular Momentum Conservation

One of the most important aspects of torque and angular momentum arises because of a curious feature of two of the most important force laws of nature: gravitation and the electrostatic force. Both of these force laws are *radial*, that is, they act ***along a line connecting two masses or charges***.

Just for grins (and to give you a quick look at them, first in a long line of glances and repetitions that will culminate in your *knowing* them, here is the simple form of the gravitational force on a "point-like" object (say, the Moon) being acted on by a second "point-like" object (say, the Earth) where for convenience we will locate the Earth at the origin of coordinates:

$$\vec{\boldsymbol{F}}_m = -G\frac{M_m M_e}{r^2}\hat{\boldsymbol{r}} \tag{6.39}$$

In this expression, $\vec{\boldsymbol{r}} = r\hat{\boldsymbol{r}}$ is the position of the moon in a spherical polar coordinate system (the direction is actually specified by two angles, neither of which affects the magnitude of the force). G is called the ***gravitational constant*** and this entire formula is a special case of ***Newton's Law of Gravitation***, currently believed to be a fundamental force law of nature on the basis of considerable evidence.

A similar expression for the force on a charged particle with charge q located at position $\vec{\boldsymbol{r}} = r\hat{\boldsymbol{r}}$ exerted a charged particle with charge Q located at the origin is known as a (special case of) ***Coulomb's Law*** and is also held to be a fundamental force law of nature. It is the force that *binds electrons to nuclei* (while making the electrons themselves repel one another) and hence is the dominant force in all of chemistry – it, more than any other force of nature, is

[149] I've seen some stupid cats land flat on their back in my lifetime, and a single counterexample serves to disprove the *absolute* rule...

"us"[150]. Coulomb's Law is just:

$$\vec{\boldsymbol{F}}_q = -k_e \frac{qQ}{r^2}\hat{\boldsymbol{r}} \tag{6.40}$$

where k_e is once again a constant of nature.

Both of these are *radial* force laws. If we compute the torque exerted by the Earth on the moon:

$$\tau_m = \vec{\boldsymbol{r}} \times \left(-G\frac{M_m M_e}{r^2}\right)\hat{\boldsymbol{r}} = 0 \tag{6.41}$$

If we compute the torque exerted by Q on q:

$$\tau_q = \vec{\boldsymbol{r}} \times \left(k_e\frac{qQ}{r^2}\right)\hat{\boldsymbol{r}} = 0 \tag{6.42}$$

Indeed, for *any* force law of the form $\vec{\boldsymbol{F}}(\vec{\boldsymbol{r}}) = F(\vec{\boldsymbol{r}})\hat{\boldsymbol{r}}$ the torque exerted by the force is:

$$\tau = \vec{\boldsymbol{r}} \times F(\vec{\boldsymbol{r}})\hat{\boldsymbol{r}} = 0 \tag{6.43}$$

and we can conclude that ***radial forces exert no torque!***

In all problems where those radial forces are the *only* (significant) forces that act:

> ***A radial force exerts no torque and the angular momentum of the object upon which the force acts is conserved.***

Note that this means that the angular momentum of the Moon in its orbit around the Earth is constant – this will have important consequences as we shall see in two or three weeks. It means that the electron orbiting the nucleus in a hydrogen atom has a constant angular momentum, at least as far as classical physics is concerned (so far). It means that if you tie a ball to a rubber band fastened to a pivot and then throw it so that the band remains stretches and shrinks as it moves around the pivot, the angular momentum of the ball is conserved. It means that when an exploding star collapses under the force of gravity to where it becomes a neutron star, a tiny fraction of its original radius, the angular momentum of the original star is (at least approximately, allowing for the mass it cast off in the *radial* explosion) conserved. It means that a mass revolving around a center on the end of a string of radius r has an angular momentum that is conserved, and that this angular momentum will *remain* conserved as the string is slowly pulled in or let out while the particle "orbits".

Let's understand this further using one or two examples.

Example 6.5.2: Mass Orbits On a String

A particle of mass m is tied to a string that passes through a hole in a frictionless table and held. The mass is given a push so that it moves in a circle of radius r at speed v. Here are several questions that might be asked – and their answers:

[150] Modulated by quantum principles, especially the notion of quantization and the Pauli Exclusion Principle, both beyond the scope of this course. Pauli is arguably co-equal with Coulomb in determining atomic and molecular structure.

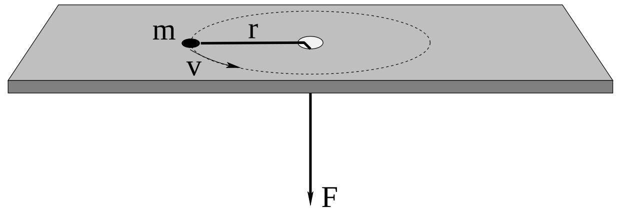


Figure 6.9

a) What is the torque exerted on the particle by the string? Will angular momentum be conserved if the string pulls the particle into "orbits" with different radii?

This is clearly a radial force – the string pulls along the vector $\vec{r}$ from the hole (pivot) to the mass. Consequently the tension in the string ***exerts no torque*** on mass m and its ***angular momentum is conserved***. It will still be conserved as the string pulls the particle in to a new "orbit".

This question is typically just asked to help *remind* you of the correct physics, and might well be omitted if this question were on, say, the final exam (by which point you are expected to have figured all of this out).

b) What is the magnitude of the angular momentum L of the particle in the direction of the axis of rotation (as a function of m, r and v)?

Trivial:

$$L = |\vec{r} \times \vec{p}| = mvr = mr^2(v/r) = mr^2\Omega = I\Omega \tag{6.44}$$

By the time you've done your homework and properly studied the examples, this should be instantaneous. Note that this is the *initial* angular momentum, and that – from the previous question – angular momentum is conserved! Bear this in mind!

c) Show that the magnitude of the force (the tension in the string) that must be exerted to keep the particle moving in a circle is:

$$F = T = \frac{L^2}{mr^3}$$

This is a *general* result for a particle moving in a circle and in no way depends on the fact that the force is being exerted by a string in particular.

As a general result, we should be able to derive it fairly easily from what we know. We know two things – the particle is moving in a circle with a constant v, so that:

$$F = \frac{mv^2}{r} \tag{6.45}$$

We also know that $L = mvr$ from the previous question! All that remains is to do some ***algebra magic*** to convert one to the other. If we had one more factor of m on to, and a factor of r^2 on top, the top would magically turn into L^2. However, we are only allowed to multiply by one, so:

$$F = \frac{mv^2}{r} \times \frac{mr^2}{mr^2} = \frac{m^2v^2r^2}{mr^3} = \frac{L^2}{mr^3} \tag{6.46}$$

as desired, Q.E.D., all done, fabulous.

d) Show that the kinetic energy of the particle in terms of its angular momentum is:

$$K = \frac{L^2}{2mr^2}$$

More straight up algebra magic of exactly the same sort:

$$K = \frac{mv^2}{2} = \frac{mv^2}{2} \times \frac{mr^2}{mr^2} = \frac{L^2}{2mr^2} \tag{6.47}$$

Now, suppose that the radius of the orbit and initial speed are r_i and v_i, respectively. From under the table, the string is *slowly* pulled down (so that the puck is always moving in an approximately circular trajectory and the tension in the string remains radial) to where the particle is moving in a circle of radius r_2.

e) Find its velocity v_2 using angular momentum conservation.

This should be very easy, and thanks to the results above, it is:

$$L_1 = mv_1r_1 = mv_2r_2 = L_2 \tag{6.48}$$

or

$$v_2 = \frac{r_1}{r_2}v_1 \tag{6.49}$$

f) Compute the work done by the force from part c) above and identify the answer *as* the work-kinetic energy theorem. Use this to to find the velocity v_2. You should get the same answer!

Well, what can we do but follow instructions. L and m are constants and we can take them right out of the integral as soon as they appear. Note that dr points out and $\vec{F}$ points in along r so that:

$$\begin{aligned} W &= -\int_{r_1}^{r_2} F dr \\ &= -\int_{r_1}^{r_2} \frac{L^2}{mr^3} dr \\ &= -\frac{L^2}{m}\int_{r_1}^{r_2} r^{-3} dr \\ &= \frac{L^2}{m}\left.\frac{r^{-2}}{2}\right|_{r_1}^{r_2} \\ &= \frac{L^2}{2mr_2^2} - \frac{L^2}{2mr_1^2} \\ &= \Delta K \end{aligned} \tag{6.50}$$

Not really so difficult after all.

Note that the last two results are pretty amazing – they show that our torque and angular momentum theory so far is remarkably *consistent* since two *very* different approaches give the same answer. Solving this problem now will make it easy later to understand the *angular momentum barrier*, the angular kinetic energy term that appears in the radial part of conservation of mechanical energy in problems involving a central force (such as gravitation and Coulomb's Law). This in turn will make it easy for us to understand certain properties of orbits from their potential energy curves.

The final application of the Law of Conservation of Angular Momentum, collisions in this text is too important to be just a subsection – it gets its very own topical section, following immediately.

6.6: Collisions

We don't need to dwell too much on the general theory of collisions at this point – all of the definitions of terms and the general methodology we learned in week 4 still hold when we allow for rotations. The primary difference is that we can now apply the Law of Conservation of Angular Momentum as well as the Law of Conservation of Linear Momentum to the actual collision impulse.

In particular, in collisions where ***no external force*** acts (in the impulse approximation), ***no external torque*** can act as well. In these collisions *both* linear momentum *and* angular momentum are conserved by the collision. Furthermore, the angular momentum can be computed relative to any pivot, so one can choose a convenient pivot to simplify the algebra involved in solving any given problem.

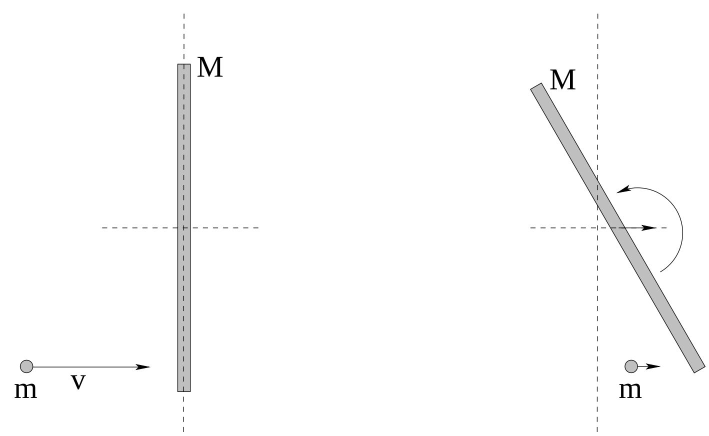


Figure 6.10: In the collision above, *no* physical pivot exist and hence *no* external force or torque is exerted during the collision. In collisions of this sort ***both momentum and angular momentum about any pivot chosen*** are conserved.

This is illustrated in figure **??** above, where a small disk collides with a bar, both sitting (we imagine) on a frictionless table so that there is no net external force or torque acting. Both momentum and angular momentum are conserved in this collision. The most convenient pivot for problems of this sort is usually the center of mass of the bar, or possibly the center of mass of the system at the instant of collision (which continues moving at the constant speed of the center of mass before the collision, of course).

All of the terminology developed to describe the *energetics* of different collisions still holds when we consider conservation of angular momentum in addition to conservation of linear momentum. Thus we can speak of *elastic* collisions where kinetic energy is conserved during the collision, and partially or fully *inelastic* collisions where it is not, with "fully inelastic" as usual being a collision where the systems collide and stick together (so that they have the same velocity of *and* angular velocity around the center of mass after the collision).

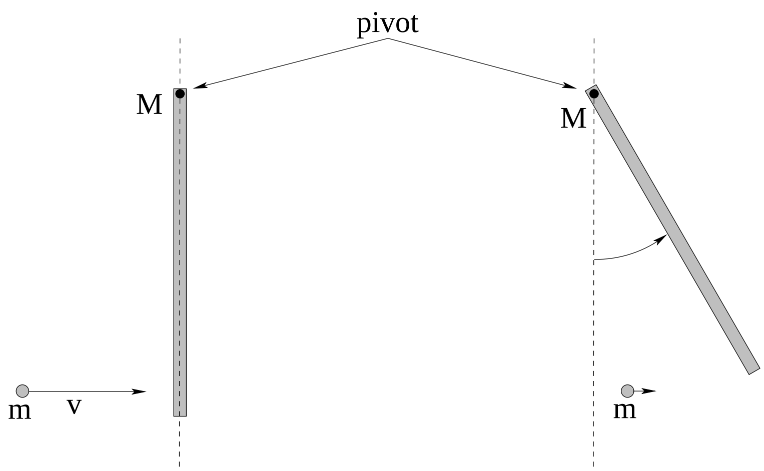


Figure 6.11: In the collision above, a physical pivot exists – the bar has a hinge at one end that prevents its linear motion while permitting the bar to swing freely. In collisions of this sort linear momentum is ***not conserved***, but since ***the pivot force exerts no torque*** about the pivot, ***angular momentum about the pivot is conserved***.

We do, however, have a new *class* of collision that can occur, illustrated in figure 6.11, one where the ***angular momentum is conserved but linear momentum is not***. This can and in general will occur when a system experiences a collision where a certain point in the system is *physically pivoted* by means of a nail, an axle, a hinge so that during the collision an *unknown force*[151] is exerted there as an extra *external* "impulse" acting on the system. This impulse acting at the pivot exerts ***no external torque*** around the pivot so ***angular momentum relative to the pivot is conserved*** but ***linear momentum is, alas, not conserved*** in these collisions.

It is extremely important for you to be able to analyze any given problem to identify the conserved quantities. To help you out, I've made up a a wee "collision type" table, where you can look for the term "elastic" in the problem – if it isn't explicitly there, by default it is at least partially inelastic unless/until proven otherwise during the solution – and also look to see if there is a *pivot force* that again by default prevents momentum from being conserved unless/until proven otherwise during the solution.

The best way to come to understand this table (and how to proceed to add angular momentum conservation to your repertoire of tricks for analyzing collisions) is by considering the following examples. I'm only doing *part* of the work of solving them here, so you can experience the joy of solving them the *rest* of the way – and learning how it all goes – for homework.

We'll start with the easiest collisions of this sort to solve – fully inelastic collisions.

[151] Often we can actually *evaluate* at least the impulse imparted by such a pivot during the collision – it is "unknown" in that it is usually not given as part of the initial data.

	Pivot Force	No Pivot Force
Elastic	K, $\vec{L}$ conserved	K, $\vec{L}$ and $\vec{P}$ conserved.
Inelastic	$\vec{L}$ conserved	$\vec{L}$ and $\vec{P}$ conserved.

Table 5: Table to help you categorize a collision problem so that you can use the correct conservation laws to try to solve it. Note that you can get *over half the credit* for any given problem simply by correctly identifying the conserved quantities even if you then completely screw up the algebra.

Example 6.6.1: Fully Inelastic Collision of Ball of Putty with a Free Rod

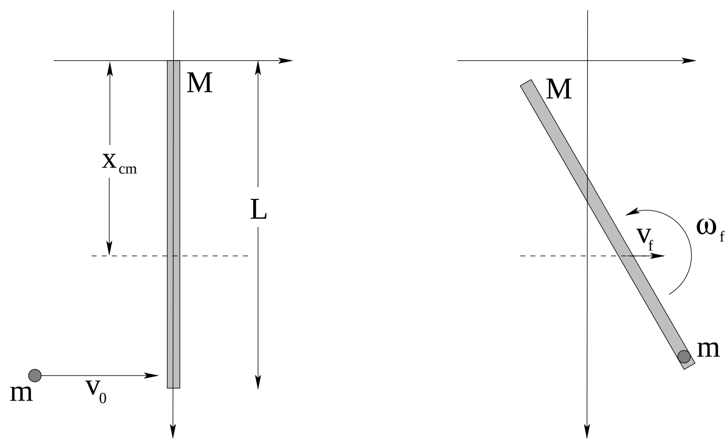


Figure 6.12: A blob of putty of mass m, travelling at initial velocity v_0 to the right, strikes an unpivoted rod of mass M and length L at the end and sticks to it. No friction or external forces act on the system.

In figure 6.12 a blob of putty of mass m strikes a stationary rod of mass M at one end and sticks. The putty and rod recoil together, rotating around their mutual center of mass. Everything is in a vacuum in a space station or on a frictionless table or something like that – in any event there are no other forces acting during the collision or we ignore them in the impulse approximation.

First we have to figure out the physics. We mentally examine our table of possible collision types. There is no pivot, so there are no relevant external forces. No external force, no external torque, so ***both momentum and angular momentum*** are conserved by this collision. However, it is a ***fully inelastic collision*** so that kinetic energy is (maximally) ***not conserved***.

Typical questions are:

- Where is the center of mass at the time of the collision (what is $x_{\rm cm}$)?
- What is v_f, the speed of the center of mass after the collision? Note that if we know the answer to these two questions, we actually know $x_{\rm cm}(t)$ for all future times!
- What is Ω_f, the final angular velocity of rotation around the center of mass? If we know this, we also know $\theta(t)$ and hence can precisely locate *every bit of mass* in the system for all times after the collision.

- How much kinetic energy is lost in the collision, and where does it go?

We'll answer these very systematically, in this order. Note well that for *each* answer, the physics knowledge required is pretty simple and well within your reach – it's just that there are a lot of parts to patiently wade through.

To find $x_{\rm cm}$:

$$(M+m)x_{\rm cm} = M\frac{L}{2} + mL \quad (6.51)$$

or

$$x_{\rm cm} = \frac{M/2+m}{M+m}L \quad (6.52)$$

where I'm taking it as "obvious" that the center of mass of the rod itself is at $L/2$.

To find v_f, we note that momentum is conserved (and also recall that the answer is *going* to be $v_{\rm cm}$:

$$p_i = mv_0 = (M+m)v_f = p_f \quad (6.53)$$

or

$$v_f = v_{\rm cm} = \frac{m}{M+m}v_0 \quad (6.54)$$

To find Ω_f, we note that *angular* momentum is going to be conserved. This is where we have to start to actually think a bit – I'm *hoping* that the previous two solutions are really easy for you at this point as we've seen each one (and worked through them in detail) at least a half dozen to a dozen times on homework and examples in class and in this book.

First of all, the good news. The rotation of ball and rod before and after the collision all happens in the plane of rotation, so we don't have to mess with anything but scalar moments of inertia and $L = I\Omega$. Then, the bad news: We have to *choose a pivot* since none was provided for us. The *answer* will be the same no matter which pivot you choose, but the algebra required to find the answer may be quite different (and more difficult for some choices).

Let's think for a bit. We know the standard scalar moment of inertia of the rod (which applies in this case) around two points – the end or the middle/center of mass. However, the final rotation is around not the center of mass of the *rod* but the center of mass of the *system*, as the center of mass of the system itself moves in a completely straight line throughout.

Of course, the angular velocity is the same regardless of our choice of pivot. We could choose the end of the rod, the center of the rod, or the center of mass of the system and in all cases the final angular momentum will be the same, but *unless we choose the center of mass of the system to be our pivot* we will have to deal with the fact that our final angular momentum will have *both* a translational *and* a rotational piece.

This suggests that our "best choice" is to choose $x_{\rm cm}$ as our pivot, eliminating the translational angular momentum altogether, and that is how we will proceed. However, I'm *also* going to solve this problem using the upper end of the rod at the instant of the collision as a pivot, because I'm quite certain that no student reading this *yet* understands what I mean about the translational component of the angular momentum!

Using $x_{\rm cm}$:

We must compute the initial angular momentum of the system before the collision. This is just the angular momentum of the incoming blob of putty at the instant of collision as the rod is at rest.

$$L_i = |\vec{r} \times \vec{p}| = mv_0 r_\perp = mv_0(L - x_{\rm cm}) \tag{6.55}$$

Note that the "moment arm" of the angular momentum of the mass m in this frame is just the perpendicular minimum distance from the pivot to the line of motion of m, $L - x_{\rm cm}$.

This must equal the final angular momentum of the system. This is easy enough to write down:

$$L_i = mv_0(L - x_{\rm cm}) = I_f \Omega_f = L_f \tag{6.56}$$

where I_f is the ***moment of inertia of the entire rotating system about the $x_{\rm cm}$ pivot!***

Note well: The ***advantage*** of using this frame with the pivot at $x_{\rm cm}$ at the instant of collision (or any other frame with the pivot on the straight line of motion of $x_{\rm cm}$) is that in this frame the ***angular momentum of the system treated as a mass at the center of mass is zero***. We *only* have a rotational part of $\vec{L}$ in any of these coordinate frames, not a rotational *and* translational part. This makes the algebra (in my opinion) ***very slightly simpler*** in this frame than in, say, the frame with a pivot at the end of the rod/origin illustrated next, although the algebra in the the frame with pivot at the origin almost instantly "corrects itself" and gives us the center of mass pivot result.

This now reveals the only point where we have to do *real work* in this frame (or any other) – finding I_f around the center of mass! Lots of opportunities to make mistakes, a need to use Our Friend, the Parallel Axis Theorem, alarums and excursions galore. However, if you have *clearly stated* $L_i = L_f$, and correctly represented them as in the equation above, you have little to fear – you might lose a point if you screw up the evaluation of I_f, you might even lose two or three, but that's out of 10 to 25 points total for the problem – you're already way up there as far as your demonstrated knowledge of physics is concerned!

So let's give it a try. The total moment of intertia is the moment of inertia of the rod around the *new* (parallel) axis through $x_{\rm cm}$ plus the moment of inertia of the blob of putty as a "point mass" stuck on at the end. Sounds like a job for the Parallel Axis Theorem!

$$I_f = \frac{1}{12}ML^2 + M(x_{\rm cm} - L/2)^2 + m(L - x_{\rm cm})^2 \tag{6.57}$$

Now be honest; this isn't really that hard to write down, is it?

Of course the "mess" occurs when we substitute this back into the conservation of momentum equation and solve for Ω_f:

$$\Omega_f = \frac{L_i}{I_f} = \frac{mv_0(L - x_{\rm cm})}{\frac{1}{12}ML^2 + M(x_{\rm cm} - L/2)^2 + m(L - x_{\rm cm})^2} \tag{6.58}$$

One could possibly square out everything in the denominator and "simplify" this, but why would one want to? If we know the *actual numerical values* of m, M, L, and v_0, we can compute (in order) $x_{\rm cm}$, I_f and Ω_f as easily from this expression as from any other, and *this* expression actually means something and can be checked at a glance by your instructors. Your instructors would have to work just as hard as you would to reduce it to minimal terms, and are just as averse to doing pointless work.

That's not to say that one should *never* multiply things out and simplify, only that it seems unreasonable to count doing so as being part of the physics of the "answer", and *all we really* *care about is the physics!* As a rule in this course, if you are a math, physics, or engineering major I expect you to go the extra mile and finish off the algebra, but if you are a life science major who came *into* the course terrified of anything involving algebra, well, I'm ***proud of you already*** because by this point in the course you have no doubt gotten *much better at math* and have started to overcome your fears – there is no need to charge you points for wading through stuff I could make a mistake doing almost as easily as you could. If anything, we'll give you *extra* points if you try it and succeed – *after* giving us something clear and correct to grade for primary credit first!

Finally, we do need to compute the kinetic energy lost in the collision:

$$\Delta K = K_f - K_i = \frac{L_f^2}{2I_f} + \frac{1}{2}(m+M)v_f^2 - \frac{1}{2}mv_0^2 \tag{6.59}$$

is as easy a form as any. *Here* there may be some point to squaring everything out to simplify, as one expects an answer that should be "some fraction of K_i", and the value of the fraction might be interesting. Again, if you are a physics major you should probably do the full simplification just for practice doing lots of tedious algebra without fear, useful self-discipline. Everybody else that does it will likely get extra credit unless the problem explicitly calls for it.

Now, let's do the whole thing *over*, using a different pivot, and see where things are the same and where they are different.

Using the end of the rod:

Obviously there is no change in the computation of $x_{\rm cm}$ and v_f – indeed, we really did these in the (given) coordinate frame starting at the end of the rod anyway – that's the "lab" frame drawn into our figure and the one wherein our answer is finally expressed. All we need to do, then, is compute our angular momenta relative to an origin/pivot at the end of the rod:

$$L_i = |\vec{\boldsymbol{r}} \times \vec{\boldsymbol{p}}| = r_\perp mv_0 = mv_0 L \tag{6.60}$$

and:

$$L_f = (m+M)v_f x_{\rm cm} + I_f\Omega_f \tag{6.61}$$

In this final expression, $(m+M)v_f x_{\rm cm}$ is the angular momentum *of* the entire system treated as a mass moving at speed v_f located at $x_{\rm cm}$ right after the collision, *plus* the angular momentum of the system *around* the center of mass, which must be computed exactly as before, same I_f, same Ω_f (to be found). Thus:

$$I_f\Omega_f = mv_0L - (m+M)v_f x_{\rm cm} \tag{6.62}$$

Here is a case where one *really must* do a bit more simplification – there are just too many things that depend on the initial conditions. If we substitute in v_f from above, in particular, we get:

$$I_f\Omega_f = mv_0L - mv_0x_{\rm cm} = mv_0(L - x_{\rm cm}) \tag{6.63}$$

and:

$$\Omega_f = \frac{mv_0(L-x_{\rm cm})}{I_f} = \frac{mv_0(L-x_{\rm cm})}{\frac{1}{12}ML^2 + M(x_{\rm cm}-L/2)^2 + m(L-x_{\rm cm})^2} \tag{6.64}$$

as before. The *algebra* somehow manages the frame change *for us*, giving us an answer that doesn't depend on the particular choice of frame once we account for the angular momentum *of* the center of mass in any frame with a pivot that isn't on the line of motion of $x_{\rm cm}$ (where it is zero). Obviously, computing ΔK is the same, and so we are done!. Same answer, two different frames!

Example 6.6.2: Fully Inelastic Collision of Ball of Putty with Pivoted Rod

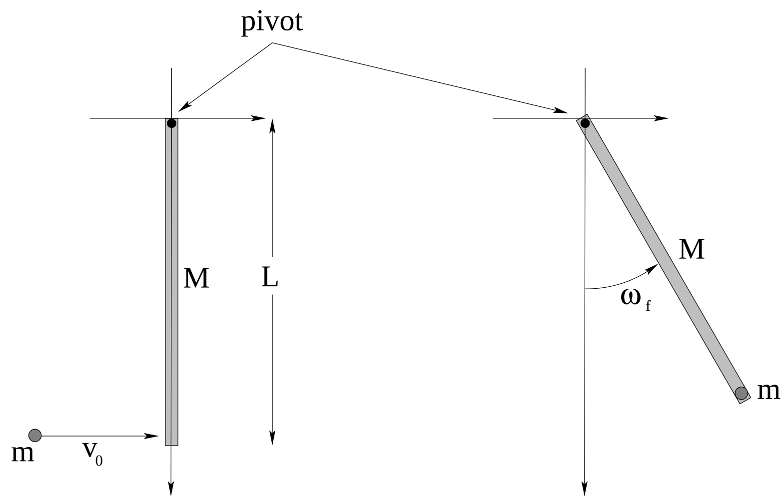


Figure 6.13: A blob of putty of mass m travelling at speed v_0 strikes the rod of mass M and length L at the end and sticks. The rod, however, is ***pivoted*** about the other end on a frictionless nail or hinge.

In figure 6.13 we see the alternative version of the fully inelastic collision between a *pivoted* rod and a blob of putty. Let us consider the answers to the questions in the previous problem. Some of them will either be exactly the same or in some sense "irrelevant" in the case of a pivoted rod, but either way we need to understand that.

First, however, the physics! Without the physics we wither and die! There is a pivot, and ***during the collision*** the pivot exerts a large and unknown pivot force on the rod[152] and this force ***cannot be ignored in the impulse approximation***. Consequently, ***linear momentum is not conserved!***

It is obviously an inelastic collision, so ***kinetic energy is not conserved***.

The force exerted by the physical hinge or nail may be unknown and large, but ***if we make the hinge the pivot*** for the purposes of computing torque and angular momentum, this force ***exerts no torque*** on the system! Consequently, for this choice of pivot *only*, ***angular momentum is conserved***.

In a sense, this makes this problem *much simpler than the last one!* We have only one physical principle to work with (plus our definitions of e.g. center of mass), and all of the other answers must be derivable from this one thing. So we might as well get to work:

$$L_i = I_0\Omega_0 = mL^2\frac{v_0}{L} = (\frac{1}{3}ML^2 + mL^2)\Omega_f = I_f\Omega_f = L_f \tag{6.65}$$

[152] Unless the mass strikes the rod at a particular point called the ***center of percussion*** of the rod, such that the velocity of the center of mass right after the collision is *exactly the same* as that of the free rod found above...

where I used $L_f = I_f \Omega_f$ and inserted the scalar moment of inertial I_f by inspection, as I know the moment of inertia of a rod about its end as well as the moment of inertia of a point mass a distance L from the pivot axis. Thus:

$$\Omega_f = \frac{mv_0 L}{(\frac{1}{3}M + m)L^2} = \frac{m}{(\frac{1}{3}M + m)}\frac{v_0}{L} = \frac{m}{(\frac{1}{3}M + m)}\Omega_0 \tag{6.66}$$

where I've written it in the latter form (in terms of the initial *angular* velocity of the blob of putty relative to this pivot) both to make the correctness of the units manifest and to illustrate how conceptually simple the answer is:

$$\Omega_f = \frac{I_0}{I_f}\Omega_0 \tag{6.67}$$

It is now straightforward to answer any other questions that might be asked. The center of mass is still a distance:

$$x_{\rm cm} = \frac{m + \frac{1}{2}M}{m + M}L \tag{6.68}$$

from the pivot, but now it moves in a circular arc after the collision, not in a straight line.

The velocity of the center of mass after the collision is determined ***from*** $\boldsymbol{\Omega_f}$***:***

$$v_f = \Omega_f x_{\rm cm} = \frac{m}{(\frac{1}{3}M + m)}\frac{v_0}{L} \times \frac{m + \frac{1}{2}M}{m + M}L = \frac{m(m + \frac{1}{2}M)}{(m + M)(\frac{1}{3}M + m)}v_0 \tag{6.69}$$

The kinetic energy lost in the collision is:

$$\Delta K = K_f - K_i = \frac{L_f^2}{2I_f} - \frac{1}{2}mv_0^2 \tag{6.70}$$

(which can be simplified, but the simplification is left as an exercise for ***everybody*** – it isn't difficult). One can even compute the ***impulse provided by the pivot hinge*** during the collision:

$$I_{\rm hinge} = \Delta p = p_f - p_i = (m + M)v_f - mv_0 \tag{6.71}$$

using v_f from above. (Note well that I have to reuse symbols such as I, in the same problem sorry – in this context it clearly means "impulse" and not "moment of inertia".)

6.6.1: More General Collisions

As you can imagine, problems can get to be somewhat more difficult than the previous two examples in several ways. For one, instead of collisions between point masses that stick and rods (pivoted or not) one can have collisions between point masses and rods where the masses do *not* stick. This doesn't change the basic physics. Either the problem will specify that the collision is elastic (so $K_{\rm tot}$ is conserved) or it will specify something like v_f for the point mass so that you can compute ΔK and possibly Δp or ΔL (depending on what *is* conserved), much as you did for bullet-passes-through-block problems in week 4.

Instead of point mass and rod *at all* you could be given a point mass and a disk or ball that can rotate, or even a collision between two disks. The algebra of things like this will be identical to the algebra above *except* that one will have to substitute e.g. the moment of inertia of a disk, or ball, or whatever for the moment of inertia of the rod, pivoted or free. The *general method* of solution will therefore be the same. You have several homework problems where you can work through this method on your own and working in your groups – make sure that you are comfortable with this before the quiz.

Angular momentum isn't just conserved in the case of symmetric objects rotating, but these are by far the easiest for us to treat. We now need to tackle the various difficulties associated with the rotation of *asymmetric* rigid objects, and then move on to finally and irrevocably understand how torque and angular momentum are ***vector quantities*** and that this *matters*.

6.7: Angular Momentum of an Asymmetric Rotating Rigid Object

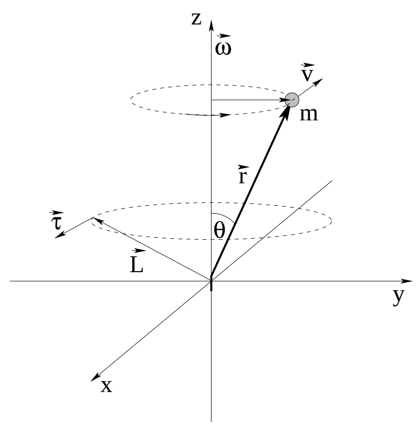


Figure 6.14: A point mass m at the end of a massless rod that makes a vector $\vec{r}$ from the origin to the location of the mass, moving at constant speed v sweeps out a circular path of radius $r\sin(\theta)$ in a plane *above* the (point) pivot. The angular momentum of this mass is (at the instant shown) $\vec{L} = \vec{r} \times m\vec{v} = \vec{r} \times \vec{p}$ as shown. As the particle sweeps out a circle, ***so does $\vec{L}$!*** The extended massless rigid rod exerts a *constantly changing/precessing torque $\vec{\tau}$* on the mass in order to accomplish this.

Consider the single particle in figure 6.14 that moves in a simple circle of radius $r\sin(\theta)$ in a plane *above the x-y plane!* The axis of rotation of the particle (the "direction of the rotation" we considered in week 5 is clearly $\vec{\Omega} = \Omega\hat{z}$.

Note well that the mass distribution of this rigid object rotation *violates both of the symmetry rules above*. It is not symmetric across the axis of rotation, nor is it symmetric across the plane of rotation. Consequently, according to our *fundamental* definition of the vector angular momentum:

$$\vec{L} = \vec{r} \times \vec{p} = \vec{r} \times m\vec{v} \tag{6.72}$$

which points *up and to the left* at the instant shown in figure 6.14.

Note well that $\vec{L}$ is ***perpendicular*** to both the plane containing $\vec{r}$ and $\vec{v}$, and that as the mass moves around in a circle, ***so does $\vec{L}$!*** In fact the vector $\vec{L}$ ***sweeps out a cone***, just as the vector $\vec{r}$ does. Finally, note that the ***magnitude*** of $\vec{L}$ has the constant value:

$$L = |\vec{r} \times \vec{v}| = mvr \tag{6.73}$$

because $\vec{r}$ and $\vec{v}$ are mutually perpendicular.

This means, of course, that although $L = |\vec{L}|$ is constant, $\vec{L}$ is constantly *changing in time*. Also, we know that the time rate of change of $\vec{L}$ is $\vec{\tau}$, so the rod must be exerting a nonzero torque on the mass! Finally, the scalar moment of inertia $I = I_{zz} = mr^2 \sin^2(\theta)$ for this rotation is a constant (and so is $\vec{\Omega}$) – manifestly $\vec{L} \neq I\vec{\Omega}$! They don't even point in the same direction!

Consider the following physics. We know that the actual magnitude and direction the force acting on m at the instant drawn is precisely $F_c = mv^2/(r\sin(\theta))$ (in towards the axis of rotation) because the mass m is *moving in a circle*. This force must be exerted by the massless rod because there is nothing else touching the mass (and we are neglecting gravity, drag, and all that). In turn, this force must be transmitted by the rod back to a bearing of some sort located at the origin, that keeps the rod from twisting out to rotate the mass in the *same* plane as the pivot (it's "natural" state of rigid rotation).

The rod exerts a torque on the mass of magnitude:

$$\tau_{\rm rod} = |\vec{r} \times \vec{F}_c| = r\cos(\theta)F_c = mv^2 \frac{\cos(\theta)}{\sin(\theta)} = m\Omega^2 r^2 \sin(\theta)\cos(\theta) \tag{6.74}$$

Now, let's see how this compares to the total change in angular momentum per unit time. Note that the *magnitude* of $\vec{L}$, L, does not change in time, nor does L_z, the component parallel to the z-axis. Only the component in the x-y plane changes, and that components sweeps out a circle!

The radius of the circle is $L_\perp = L\cos(\theta)$ (from examining the various right triangles in the figure) and hence the total change in $\vec{L}$ in one revolution is:

$$\Delta L = 2\pi L_\perp = 2\pi L \cos(\theta) \tag{6.75}$$

This change occurs in a time interval $\Delta t = T$, the period of rotation of the mass m. The period of rotation of m is the distance it travels (circumference of the circle of motion) divided by its speed:

$$T = \frac{2\pi r \sin(\theta)}{v} \tag{6.76}$$

Thus the magnitude of the torque exerted over $\Delta t = T$ is (using $L = mvr$ as well):

$$\frac{\Delta L}{\Delta t} = 2\pi L\cos(\theta)\frac{v}{2\pi r\sin(\theta)} = mv^2\frac{\cos(\theta)}{\sin(\theta)} = m\Omega^2 r^2 \sin(\theta)\cos(\theta) \tag{6.77}$$

so that indeed,

$$\vec{\tau}_{\rm rod} = \frac{\Delta L}{\Delta t} \tag{6.78}$$

for the cycle of motion.

The rod must indeed exert a *constantly changing* torque on the rod, a torque that ***remains perpendicular*** to the angular momentum vector at all times. The particle itself, the angular momentum, and the torque acting on the angular momentum all ***precess*** around the z-axis with a period of revolution of T and clearly at no time is it true that $\vec{L} = I\vec{\Omega}$ for any scalar I and constant $\vec{\Omega}$.

Notice how things change if we *balance* the mass with a *second* one on an opposing rod as drawn in figure 6.15, making the distribution ***mirror symmetric across the axis of rotation***.

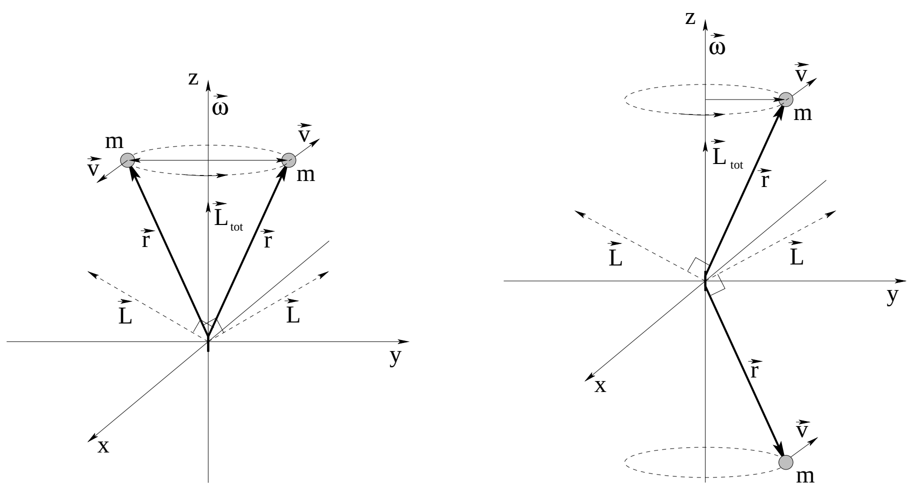


Figure 6.15: When the two masses have ***mirror symmetry across the axis of rotation***, their *total* angular momentum $\vec{L}$ *does* line up with $\vec{\Omega}$. When the two masses have ***mirror symmetry across the plane of rotation***, their *total* angular momentum $\vec{L}$ *also* lines up with $\vec{\Omega}$.

Now each of the two masses has a torque acting on it due to the rod connecting it with the origin, each mass has a vector angular momentum that at right angles to both $\vec{r}$ and $\vec{v}$, but the components of $\vec{L}$ in the x-y plane phcancel so that the total angular momentum once again lines up with the z-axis!

The same thing happens if we add a second mass at the mirror-symmetric position ***below*** the plane of rotation as shown in the second panel of figure **??**. In this case as well the components of $\vec{L}$ in the x-y plane cancel while the z components add, producing a total vector angular momentum that points in the z-direction, parallel to $\vec{\Omega}$.

The two ways of balancing a mass point around a pivot are not quite equivalent. In the first case, the pivot axis passes through the center of mass of the system, and the rotation can be maintained without any external *force* as well as torque. In the second case, however, the center of mass of the system itself is moving in a circle (in the plane of rotation). Consequently, while no net external *torque* is required to maintain the motion, there is a net external *force* required to maintain the motion. We will differentiate between these two cases below in an everyday example where they matter.

This is why I at least *tried* to be careful to assert throughout week 5) that the mass distributions for the one dimensional rotations we considered were sufficiently *symmetric*. "Sufficiently", as you should now be able to see and understand, means mirror symmetric across the axis of rotation (best, zero external force or torque required to maintain rotation)) or plane of rotation (sufficient, but need external force to maintain motion of the center of mass in a circle). Only in these two cases is the total angular momentum $\vec{L}$ is parallel to $\vec{\Omega}$ such that $\vec{L} = I\vec{\Omega}$ for a suitable scalar I.

Example 6.7.1: Rotating Your Tires

This is why you should regularly *rotate your tires* and keep them *well-balanced*. A "perfect tire" is one that is precisely cylindrically symmetric. If we view it from the side it has a uniform mass

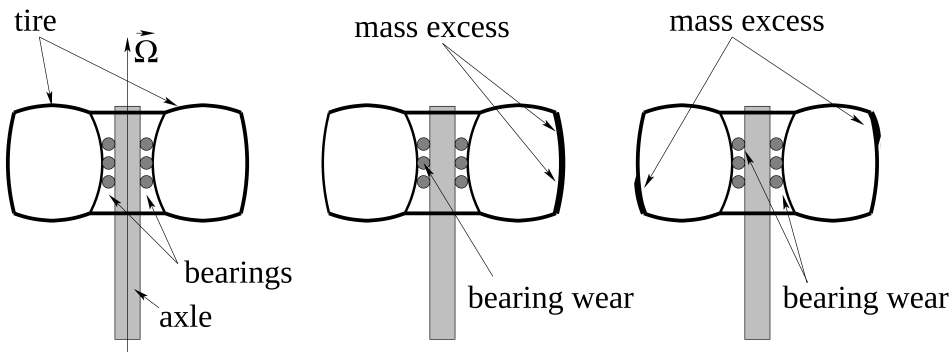


Figure 6.16: Three tires viewed in cross section. The first one is perfectly symmetric and balanced. The second one has a static imbalance – one side is literally more massive than the other. It will stress the bearings as it rotates as the bearings have to exert a differential centripetal force on the more massive side. The third is dynamically imbalanced – it has a non-planar mass assymetry and the bearings will have to exert a constantly precessing *torque* on the tire. Both of the latter situations will make the drive train noisy, the car more difficult and dangerous to drive, and will wear your bearings and tires out much faster.

distribution that has ***both*** mirror symmetry across the axis of rotation ***and*** mirror symmetry across the plane of rotation. If we mount such a tire on a frictionless bearing, no particular side will be heavier than any other and therefore be more likely to rotate down towards the ground. If we spin it on a frictionless axis, it will spin perfectly symmetrically as the bearings will not have to exert any ***precessing torque*** or ***time-varying force*** on it of the sort exerted by the massless rod in figure 6.14.

For a variety of reasons – uneven wear, manufacturing variations, accidents of the road – tires (and the hubs they are mounted on) rarely stay in such a perfect state for the lifetime of either tire or car. Two particular kinds of imbalance can occur. In figure 6.16 three tires are viewed in cross-section. The first is our mythical brand new perfect tire, one that is both statically and dynamically balanced.

The second is a tire that is *statically* imbalanced – it has mirror symmetry across the plane of rotation but not across the axis of rotation. One side has thicker tread than the other (and would tend to rotate down if the tire were elevated and allowed to spin freely). When a statically imbalanced tire rotates while driving, the center of mass of the tire moves in a circle around the axle and the bearings on the *opposite* side from the increased mass are anomalously compressed in order to provide the required centripetal force. The car bearings will wear faster than they should, and the car will have an annoying vibration and make a wubba-wubba noise as you drive (the latter can occur for many reasons but this is one of them).

The third is a tire that is *dynamically* imbalanced. The surplus masses shown are balanced well enough from left to right – neither side would roll down as in static imbalance, but the mass distribution does *not* have mirror symmetry across the axis of rotation *nor* does it have mirror symmetry across the plane of rotation through the pivot. Like the unbalanced mass sweeping out a cone, the bearings have to exert a *dynamically changing torque* on the tires as they rotate because their angular momentum is *not parallel to the axis of rotation*. At the instant shown, the bearings are stressed in *two* places (to exert a net torque on the hub *out of the page* if the direction of $\vec{\Omega}$ is up as drawn).

The solution to the problem of tire imbalance is simple – rotate your tires (to maintain even wear) and have your tires regularly *balanced* by adding compensatory weights on the "light" side of a static imbalance and to restore relative cylindrical symmetry for dynamical imbalance, the way adding a second mass did to our single mass sweeping out a cone.

I should point out that there are *other* ways to balance a rotating rigid object, and that *every* rigid object, no matter how its mass is distributed, has at least certain axes through the center of mass that can "diagonalize" the moment of inertia with respect to those axes. These are the axes that you can spin the object about and it will rotate freely without any application of an outside force or torque. Sadly, though, this is beyond the scope of this course[153]

At this point you should understand how easy it is to evaluate the angular momentum of symmetric rotating rigid objects (given their moment of inertia) using $L = I\Omega$, and how very difficult it can be to manage the angular momentum in the cases where the mass distribution is asymmetric and unbalanced. In the latter case we will usually find that the angular momentum vector will *precess* around the axis of rotation, necessitating the application of a *continuous torque* to maintain the (somewhat "unnatural") motion.

There is one other very important context where precession occurs, and that is when a *symmetric* rigid body is rapidly rotating and has a large angular momentum, and a torque is applied to it in just the right way. This is one of the most important problems we will learn to solve this week, one essential for *everybody* to know, future physicians, physicists, mathematicians, engineers: The Precession of a Top (or other symmetric rotator).

[153]Yes, I know, I know – this was a *joke!* I know that you aren't, in fact, terribly sad about this...;-)

6.8: Precession of a Top

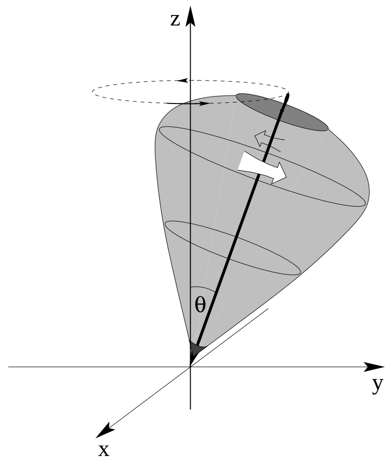


Figure 6.17: A spinning top precesses around the vertical axis as long as its angular momentum along its axis is much larger than any associated with the motion of its center of mass relative to the pivot.

Nowhere is the vector character of torque and angular momentum more clearly demonstrated than in the phenomenon of ***precession*** of a top, a spinning proton, or the Earth itself. This is also an *important* problem to clearly and quantitatively understand even in this introductory course because it is ***the basis of Magnetic Resonance Imaging (MRI)*** in medicine and biology, the basis of understanding quantum phenomena ranging from (nuclear or electron) ***spin resonance*** to ***resonant emission from two-level atoms*** for physicists, and the basis for ***gyroscopes*** to the engineer. Everybody needs to know it, in other words, no fair hiding behind the sputtered "but I don't need to know this crap" weasel-squirm all too often uttered by frustrated students.

It is therefore worth your while to invest the time required to *completely understand* precession in terms of vector torque, and to be able to ***derive*** the angular frequency of precession, Ω_p and ***predict*** the direction of precession on a quiz or exam question. I'll show you three ways to do it below, and any of the three will be acceptable (although I prefer that physics majors and perhaps engineers use the second or third as the first is a bit *too* simple; it gets the right answer but doesn't give you a good feel for what happens if the forces that produce the torque change in time).

First, though, let's understand the phenonenon. In figure 6.17 above, a simple top is shown spinning around its axis of symmetry at some angular velocity $\vec{\boldsymbol{\Omega}}$ (not to be confused with Ω_p, the angular velocity of precession, that we would like to derive).

We begin by noting that this top is symmetric, so we can easily compute the magnitude and direction of its ***angular momentum***:

$$L = |\vec{\boldsymbol{L}}| = I\Omega = \beta M R^2 \Omega \tag{6.79}$$

where β can be adjusted to make the solution apply to gyroscopic rings, disks, balls, etc. as well as tops.

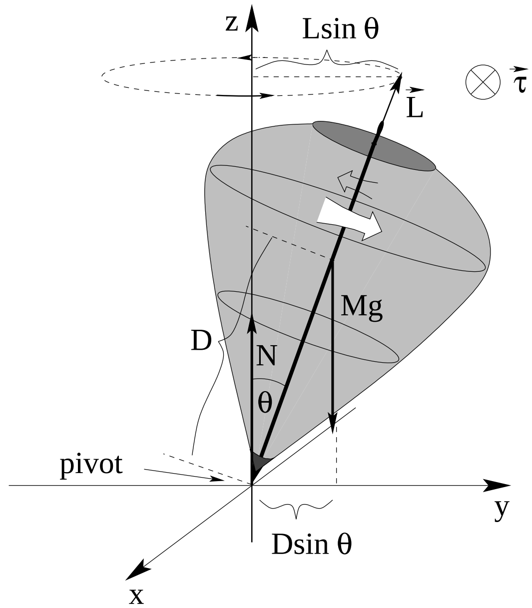


Figure 6.18: The top redrawn with forces and torques and angular momentum added.

Using the "grasp the axis" version of the right hand rule in figure 6.18, we see that in the example portrayed this angular momentum points *in the direction from the pivot at the point of contact between the spindle and the ground and the center of mass of the disk* along the axis of rotation.

Second, we note that there is a ***net torque*** exerted on the top relative to this spindle-ground contact point pivot due to gravity. There is no torque, of course, from the normal force that holds up the top as long as we locate the pivot there, and we are assuming the spindle is held at a fixed point on the ground but exerts no frictional torque where it touches the ground. The gravitational torque is a *vector*:

$$\vec{\tau} = \vec{D} \times (-Mg\hat{z}) \tag{6.80}$$

so the vector torque is given by:

$$\tau = MgD\sin\theta \tag{6.81}$$

with a direction ***into the page*** (negative x) at the instant drawn above, where z is vertical and $\vec{L}$ has z and y components only.

Third we note that the torque will change the angular momentum by displacing it *into the page* in the direction of the torque. But as it does so, the plane containing the $\vec{D}$ and $Mg\hat{z}$ will be rotated a tiny bit around the z (precession) axis, and the torque will *also* shift its direction to *remain* perpendicular to this plane. It will shift $\vec{L}$ a bit more, which shifts $\vec{\tau}$ a bit more and so on. In the end $\vec{L}$ will ***precess around*** z, with $\vec{\tau}$ precessing as well, always $\pi/2$ ahead .

Since $\vec{\tau} \perp \vec{L}$, the magnitude of $\vec{L}$ does not change, only the direction. Heuristically, then, $\vec{L}$ *sweeps out a cone* much like the cone swept out in the unbalanced rotation problems above. We can use similar considerations to relate the magnitude of the torque (known above) to the total cumulative change of angular momentum over a period. This is the simplest (and least rigorous) way to find the precession frequency. Let's start by giving this a try:

Example 6.8.1: Finding Ω_p From $\Delta L/\Delta t$ (Average)

We already showed above that

$$\tau_{\rm avg} = \tau = MgD\sin\theta \tag{6.82}$$

is the *magnitude* of the torque at all points in the precession cycle, and hence is also the average of the magnitude of the torque over a precession cycle (as opposed to the average of the vector torque over a cycle, which is obviously zero).

We now *also* compute the average torque algebraically from the cumulative value of $\Delta L/\Delta t$ over a single precession cycle – also a kind of average – and equate the two forms. That is:

$$\Delta L_{\rm cycle} = 2\pi L_\perp = 2\pi L\sin\theta \tag{6.83}$$

and:

$$\Delta t_{\rm cycle} = T_p = \frac{2\pi}{\Omega_p} \tag{6.84}$$

where Ω_p is the (desired) precession frequency, or:

$$\tau_{\rm avg} = MgD\sin\theta = \frac{\Delta L_{\rm cycle}}{\Delta t_{\rm cycle}} = \left(\frac{\Omega_p}{2\pi}\right)\times(2\pi L\sin\theta) = \Omega_p L\sin\theta \tag{6.85}$$

We now solve for the precession frequency Ω_p:

$$\Omega_p = \frac{MgD\cancel{\sin\theta}}{L\cancel{\sin\theta}} = \frac{MgD}{L} = \frac{MgD}{I\Omega} = \frac{gD}{\beta R^2\Omega} \tag{6.86}$$

or

$$\boxed{\Omega_p = \frac{gD}{\beta R^2\Omega}} \tag{6.87}$$

This result is reasonably general and can easily be adapted/computed for tops of various sizes and shapes.

Note well that the precession frequency is ***independent of θ*** – this is extremely important next semester when you study the precession of spinning charged particles around an applied magnetic field, the basis of MRI!

Example 6.8.2: Finding Ω_p from ΔL and Δt Separately

Simple and general as it is, the previous derivation has a few "issues" and hence it is not my favorite one; averaging in this way doesn't give you the most insight into what's going on and doesn't help you get a good feel for the calculus of it all. It is adequate, perhaps, for gyroscopes and tops, but I'd also like to prepare you for a future Electricity and Magnetism course where solving similar equations of motion is crucial to understanding nuclear magnetic resonance (NMR), electron spin resonance (ESR), and MRI. We'll therefore do two more derivations that improve on the crude averaging up above, working towards a rigorous application of calculus to obtain the coupled differential equations of the motion for the problem that we won't be able to *completely* understand and solve until we reach the chapter on oscillation.

First, let's repeat the general idea of the previous derivation, only instead of setting the *magnitude* τ equal to the *average, cumulated* magnitude of the change in $\vec{L}$ over an entire precession period T_p, we will set $\vec{\tau}$ equal to the *instantaneous* change in $\vec{L}$ per unit time, writing everything in terms of very small (but finite) intervals and then taking the appropriate limits to make them infinitesimal.

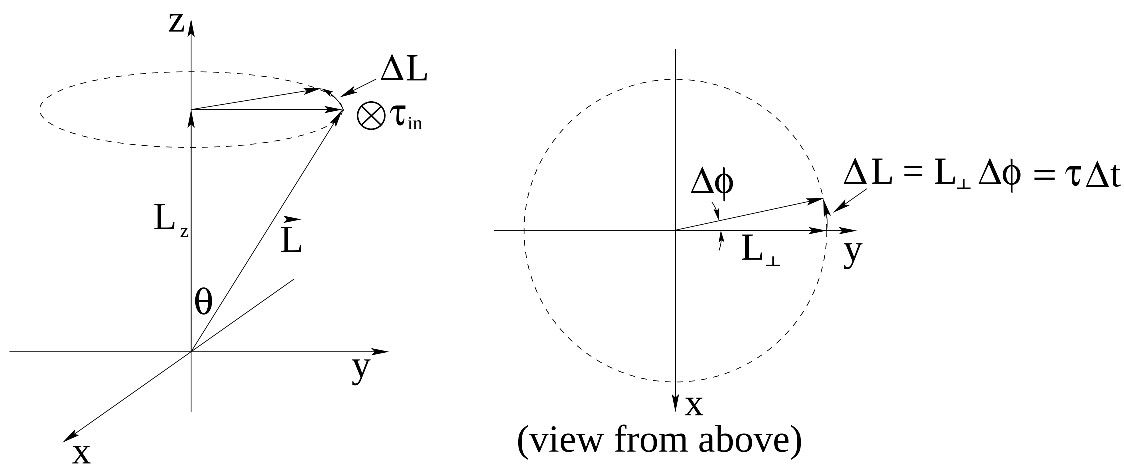


Figure 6.19: The cone swept out by the precession of the angular momentum vector, $\vec{L}$, as well as an "overhead view" of the trajectory of $\vec{L}_\perp$, the component of $\vec{L}$ perpendicular to the z-axis.

To do this, we need to *picture* how $\vec{L}$ changes in time without the distraction of drawing the entire top and forces. We'll start (in figure 6.19) with $\vec{L}$ in the y-z plane, with a positive y component at $t = 0$. At any given instant $\vec{\tau}$ is perpendicular to *both* $\vec{L}$ and $\vec{F}_g = -Mg\hat{z}$ and ***cannot change the magnitude of $\vec{L}$ or its z component, L_z***!

Over a very short time Δt, the change in $\vec{L}$ is thus $\Delta\vec{L}$ in the plane perpendicular to $\hat{z}$. As the instant drawn above, the angular momentum vector changes *direction only* in the $\vec{\tau}$ (or $-\hat{x}$) direction. At the same time, the $\vec{\tau}$ direction *also* also changes slightly to *remain* perpendicular. This should remind you of our discussion in chapter/week one of the kinematics of circular motion.

Here's a rigorous argument for $\vec{L}$ sweeping out a cone. As we just observed, $L_z\hat{z}$ must remain constant. The magnitude of $\vec{L}$ (squared) $L^2 = (L_x^2 + L_y^2) + L_z^2$ must also remain constant. If we define the *magnitude* of the part of $\vec{L}$ perpendicular to $\hat{z}$ to be $L_\perp = +\sqrt{L_x^2 + L_y^2}$, then *this* must remain constant (although its direction does not have to be constant and indeed must change!) We can now conclude that $\vec{L}$ sweeps out a *cone* around the z-axis where L_z and the magnitude of $L_\perp$ remain constant.

In other words, ***$\vec{L}_\perp$ must sweep out a circle over time!*** The change in the angular momentum $\vec{L}$ over the *short* (soon to be differential) time Δt is the *directed circular arc* $\Delta\vec{L}$ subtended by the *small* angle $\Delta\phi$ portrayed in figure 6.19 through which it precesses in this time. We can see from the top view that its magnitude is given by:

$$\Delta L = L_\perp \Delta\phi = \tau\Delta t \tag{6.88}$$

where $\Delta\phi$ is the angle the angular momentum vector precesses through in time Δt. We substitute $L_\perp = L\sin\theta$ and $\tau = MgD\sin\theta$ as before, and get:

$$L\cancel{\sin\theta}\,\Delta\phi = MgD\cancel{\sin\theta}\,\Delta t \tag{6.89}$$

We cancel $\sin\theta$ as shown and solve for:

$$\Omega_p = \frac{d\phi}{dt} = \lim_{\Delta t \to 0} \frac{\Delta\phi}{\Delta t} = \frac{MgD}{L} = \frac{gD}{\beta R^2 \Omega} \tag{6.90}$$

as before.

Because *this* time we could take the limit as $\Delta t \to 0$, we get an expression that is good at *any* instant in time, even if the top is in a rocket ship (an accelerating frame) and $g \to g'$ is varying in time! However, this *still* isn't the most elegant, or precise approach as it doesn't lead us directly to $\vec{\boldsymbol{L}}(t)$, the actual angular momentum as a function of time. The *best* approach, although it *does* use some real calculus, is to just write down the equation of motion for the system as differential equations and solve them for *both* $\vec{\boldsymbol{L}}(t)$ *and* get Ω_p and the direction directly, without any (right!) hand waving to get magnitudes or directions.

Example 6.8.3: Finding Ω_p from Calculus

Way back at the beginning of this section we wrote down Newton's Second Law for the rotation of the gyroscope *directly*:

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{D}} \times (-Mg\hat{\boldsymbol{z}}) \tag{6.91}$$

Because $-Mg\hat{\boldsymbol{z}}$ points only in the negative z-direction and

$$\vec{\boldsymbol{D}} = D\frac{\vec{\boldsymbol{L}}}{L} \tag{6.92}$$

because $\vec{\boldsymbol{L}}$ is parallel to $\vec{\boldsymbol{D}}$, for a general $\vec{\boldsymbol{L}} = L_x\hat{\boldsymbol{x}} + L_y\hat{\boldsymbol{y}}$ we get precisely *two terms* out of the cross product:

$$\frac{dL_x}{dt} = \tau_x = \frac{MgDL_y}{L} \tag{6.93}$$

$$\frac{dL_y}{dt} = \tau_y = -\frac{MgDL_x}{L} \tag{6.94}$$

If we differentiate the first expression and substitute the second into the first (and vice versa) we transform this pair of coupled *first* order differential equations for L_x and L_y into the following pair of *second* order differential equations:

$$\frac{d^2L_x}{dt^2} = -\left(\frac{MgD}{L}\right)^2 L_x = -\Omega_p^2 L_x \tag{6.95}$$

$$\frac{d^2L_y}{dt^2} = -\left(\frac{MgD}{L}\right)^2 L_y = -\Omega_p^2 L_y \tag{6.96}$$

These (either one) we will learn to recognize as the differential equation of motion for the *simple harmonic oscillator*. A particular solution of interest (that satisfies the first order equations above as well as the given initial condition that $\vec{\boldsymbol{L}}(0) = L_\perp\hat{\boldsymbol{y}} + L_{z0}\hat{\boldsymbol{z}}$) might be:

$$L_x(t) = -L_\perp \sin(\Omega_p t) \tag{6.97}$$

$$L_y(t) = +L_\perp \cos(\Omega_p t) \tag{6.98}$$

$$L_z(t) = L_{z0} \tag{6.99}$$

where $L_\perp$ is the magnitude of the component of $\vec{\boldsymbol{L}}$ which is perpendicular to $\hat{z}$. This is then the *exact solution* to the equation of motion for $\vec{\boldsymbol{L}}(t)$ that describes the actual cone being swept out from the given initial conditions, with no hand-waving or limit taking required.

The only catch to this approach is, of course, that we don't know how to solve the equation of motion yet, and the very *phrase* "second order differential equation" strikes terror into our hearts ***in spite of the fact that we've been solving one after another since week 1 in this class! All*** of the equations of motion we have solved from Newton's Second Law have been second order ones, after all – it is just that the ones for constant acceleration were directly *integrable* where this set is not, at least not easily.

It is pretty easy to solve for all of that, but we will postpone the actual solution until later. In the meantime, remember, you must know how to reproduce *one of the three derivations above* for Ω_p, the angular precession frequency, for *at least one quiz, test, or hour exam*. Not to mention a homework problem, below. Be sure that you master this because precession is *important*.

One last suggestion before we move on to treat angular collisions. Most students have a lot of experience with pushes and pulls, and so far it has been pretty easy to come up with everyday experiences of force, energy, one dimensional torque and rolling, circular motion, and all that. It's not so easy to come up with everyday experiences involving vector torque and precession. Yes, may of you played with tops when you were kids, yes, nearly everybody *rode bicycles* and balancing and steering a bike involves torque, but you haven't really *felt* it knowing what was going on.

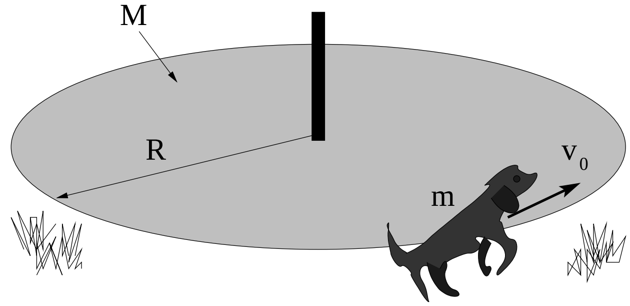
The only device likely to help you to personally experience torque "with your own two hands" is the bicycle wheel with handles and the string that was demonstrated in class. I ***urge you to take a turn spinning this wheel*** and trying to turn it by means of its handles while it is spinning, to spin it and suspend it by the rope attached to one handle and watch it precess. Get to where you can predict the direction of precession given the direction of the spin and the handle the rope is attached to, get so you have *experience* of pulling it (say) in and out and feeling the handles deflect up and down or vice versa. Only thus can you ***feel*** the nasty old cross product in both the torque and the angular momentum, and only thus can you come to understand torque with your gut as well as your head.

Homework for Week 6

Problem 1.

Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

Problem 2.

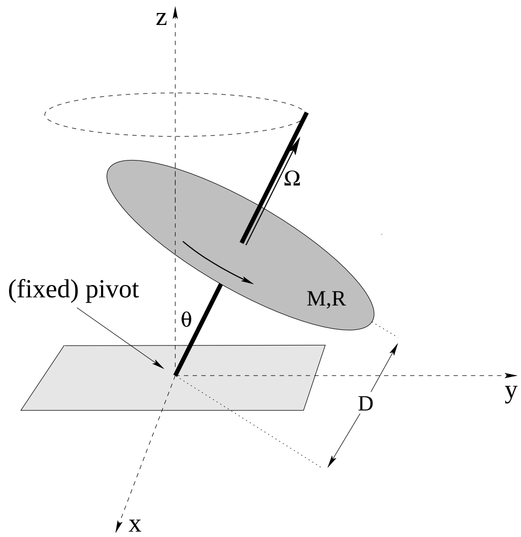


Satchmo (a dog with mass m) runs and jumps onto the edge of a merry-go-round (that is initially at rest) and then sits down there to take a ride as it spins. The merry-go-round can be thought of as a disk of radius R and mass M, and has approximately frictionless bearings in its axle. At the time of this angular "collision" Satchmo is travelling at speed v_0 perpendicular to the radius of the merry-go-round and you can neglect Satchmo's moment of inertia about an axis through his OWN center of mass compared to that of Satchmo travelling around the merry-go-round axis (because R of the merry-go-round is much larger than Satchmo's size, so we can treat him like "a particle").

a) What is the angular velocity Ω_f of the merry-go-round (and Satchmo) right after the collision?

b) How much of Satchmo's initial kinetic energy was lost, compared to the final kinetic energy of Satchmo plus the merry-go-round, in the collision? (Believe me, no matter what he's lost it isn't enough – Satchmo is – or rather was, sadly – a border collie and he had plenty more!)

Problem 3.



A top is made of a uniform ***disk*** of radius R and mass M with a very thin, light (assume massless) nail for a spindle so that the center of the disk is a distance D from the tip, which is caught in a small frictionless depression so that it acts as a *fixed pivot*. The top is spun with a large angular velocity Ω with the nail vertically above the y-axis as shown above. Recall that the free motion of the top is then *precession*.

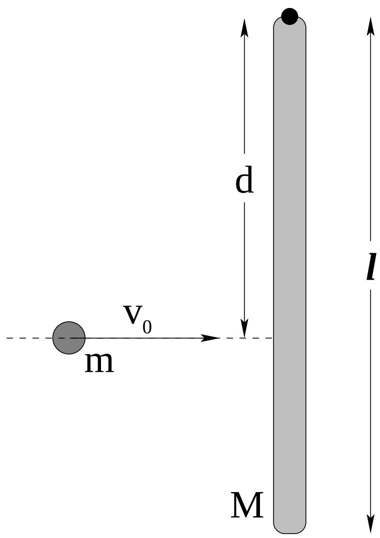
a) Find the ***vector*** torque $\vec{\tau}$ exerted about the pivot at the instant shown in the figure. You may express the vector any (correct) way that you wish (e.g. magnitude and direction, cartesian components).

b) What is the axis of precession (the vector direction of $\vec{\Omega}_p$)?

c) Derive the angular velocity of precession $\vec{\Omega}_p$. Any of the derivations used in class or discussed in the textbook are acceptable.

Express all answers in terms of M, R, g, D, and θ as needed.

Problem 4.



A rod of mass M and length ℓ rests on a frictionless table and is pivoted on a frictionless nail at one end as shown. A blob of putty of mass m approaches with velocity v_0 from the left and strikes the rod a distance d from the end as shown, sticking to the rod.

a) Find the final angular velocity Ω_f of the system about the nail after the collision.

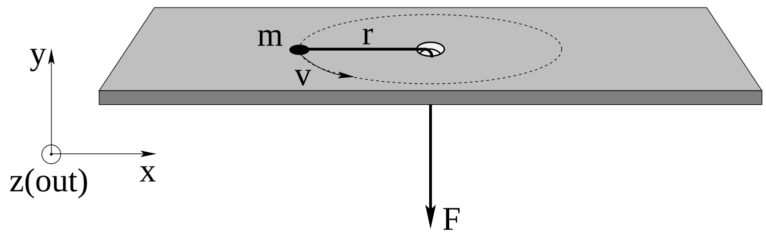
b) Is the linear momentum of the rod/blob system conserved in this collision for a ***general*** value of d? If not, why not?

c) Find the ***specific*** value d_c for which it *is* conserved? This value is called the ***c****enter of percussion* of the rod.

All answers should be in terms of M, m, ℓ, v and d as needed. Note well that (as always) you should clearly indicate what physical principles you are using to solve this problem. You may also find it useful to read:

Wikipedia: http://www.wikipedia.org/wiki/Center of Percussion

Problem 5.



A particle of mass M is tied to a string that passes through a hole in a frictionless table and held. The mass is given a push so that it initially moves in a circle of radius r_i at speed v_i. We will now conceptual review and algebraically analyze the physics of its motion in two stages. Please answer the following questions. While the string is fixed (so that r_i is constant):

a) What is the torque exerted on the particle by the string?

b) What is the ***vector angular momentum*** L_i of the particle? Use the provided coordinate system to give the direction.

c) Show that the magnitude of the force (the tension in the string) that must be exerted to keep the particle moving in this circle is:

$$F = \frac{L_i^2}{mr_i^3}$$

Note that ***this is a general result*** for a particle moving in a circle and in no way depends on the fact that the force is being exerted by a string in particular.

d) Show that the kinetic energy of the particle in terms of its angular momentum is:

$$K_i = \frac{L_i^2}{2mr_i^2}$$

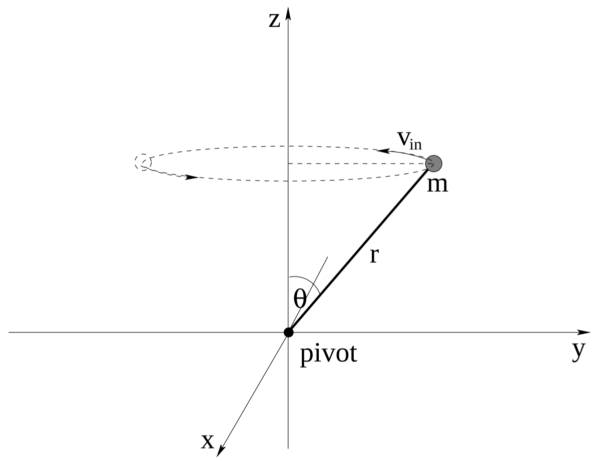
From under the table, the string is *slowly* pulled down (so that the puck is always moving in an approximately circular trajectory and the tension in the string remains radial) to where the particle is moving in a circle of radius r_f.

e) If the tension in the string remains radial, what quantity ought to be conserved?

f) Find its velocity v_f using conservation of that quantity.

g) Compute the work done by the force from part c) above and identify the answer *as* the work-kinetic energy theorem. Use this principle *instead* to to find the velocity v_f. You should get the same answer!

Problem 6.



In the figure above, a mass m is being spun in a circular path at a constant speed v around the z-axis on the end of a massless rigid rod of length r pivoted at the orgin that rides on bearings at the origin so that it sweeps out a right circular cone at an angle θ around the z axis as shown. All answers should be given in terms of m, r, θ and v. Ignore gravity, friction, and drag.

a) Find the angular momentum vector of this system $\vec{L}$ *relative to the pivot* at the instant shown and draw it in on (a copy of) the figure.

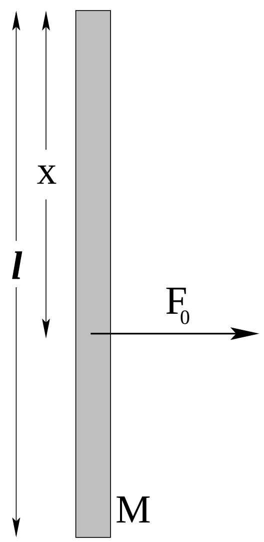
b) Find the angular velocity vector of the mass m at the instant shown and draw *it* in on the figure. Is $\vec{L} = I_z\vec{\Omega}$, where $I_z = mr^2\sin^2(\theta)$ is the moment of inertia of a mass m circling around the z axis? What *component* of the angular momentum is equal to $I_z\Omega$?

c) What must the magnitude and direction of the torque exerted by the rod on the mass (relative to the pivot shown) be in order to explain the change in angular momentum as the mass circulates at constant speed? (Hint: What is the rate that the angular momentum changes as it sweeps out a cone of its own?)

d) What is the *direction* of the force exerted by the rod in order for the mass to move in a circle to create this torque? (Hint: Shift gears and think about the actual trajectory of the particle. What must the direction of the force be?)

e) Set $\tau = rF_\perp$ or $\tau = r_\perp F$ and determine the magnitude of this torque. Note that it is equal in magnitude to your answer to c) above and has just the right direction!

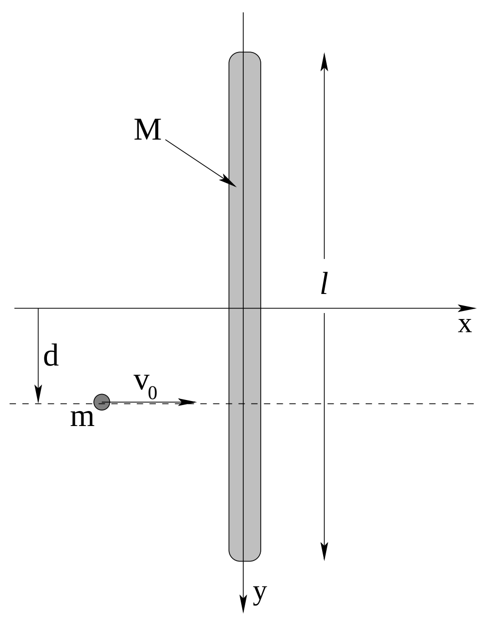
Advanced Problem 7.



A rod of mass M and length ℓ rests on a frictionless table. An ***impulse force*** with *average magnitude* F_0 is exerted on the rod a distance x from the end of the rod for a very short time Δt.

Prove that ***the top end of the rod will remain stationary*** (as if it were "pivoted" there with a pivot that exerted no force on the system during the collision) when $x = 2\ell/3$. Based on your work on a regular homework problem, how can we interpret this result?

Advanced Problem 8.



A rod of mass M and length L rests on a frictionless table. A blob of putty of mass m approaches with velocity v from the left and strikes the rod a distance d from the end as shown, sticking to the rod.

a) Find the final angular velocity $\vec{\Omega}_f$ of the system about the center of mass ***of the system*** after the collision. (Note that the rod and putty will not be rotating about the center of mass of the rod, and don't forget direction!)

b) Will the linear momentum of the rod/blob system conserved in this collision for any value of d? If so, why? If not, why not?

c) Is kinetic energy conserved in this collision? If not, how much is lost? Where does the energy go?

All answers should be in terms of M, m, L, v and d as needed.

Week 7: Statics

1.13: Statics Summary

- A rigid object is in ***static equilibrium*** when ***both*** the ***vector torque and the vector force*** acting on it are ***zero***. That is:

  If $\vec{F}_{\rm tot} = 0$ and $\vec{\tau}_{\rm tot} = 0$, then an object initially at rest will remain at rest and neither accelerate nor rotate.

  This rule applies to particles with intrinsic spin as well as "rigid objects", but this week we will primarily concern ourselves with rigid objects as static *force* equilibrium for particles was previously discussed.

  Note well that the torque and force in the previous problem are both ***vectors***. In many problems they will effectively be one dimensional, but in some they will ***not*** and you must establish e.g. the torque equilibrium condition for several different directions.

- A common question that arises in statics is the ***tipping problem***. For an object placed on a slope or pivoted in some way such that *gravity opposed by normal forces* provides one of the sources of torque that tends to keep the object *stable*, while some variable force provides a torque that tends to *tip the object over* the pivot, one uses the condition of *marginal static equilibrium* to determine, e.g. the lowest value of the variable force that will tip the object over.

- A force ***couple*** is defined to be a pair of forces that are equal and opposite but that ***do not necessarily or generally act along the same line upon an object***. The point of this definition is that it is easily to see that force couples ***exert no net force*** on an object but they ***will exert a net torque on the object*** as long as they do *not* act along the same line. Furthermore:

- The vector torque exerted on a rigid object by a force couple is the ***same for all choices of pivot!*** This (and the frequency with which they occur in problems) is the basis for the definition.

As you can see, this is a short week, just perfect to share with the midterm hour exam.

7.1: Conditions for Static Equilibrium

We already know ***well*** (I hope) from our work in the first few weeks of the course that an object at rest remains at rest *unless acted on by a net external force!* After all, this is just Newton's First Law! If a particle is located at some position in any inertial reference frame, and isn't moving, it won't *start* to move unless we push on it with some force produced by a law of nature.

Newton's Second Law, of course, applies only to *particles* – infinitesimal chunklets of mass in extended objects or elementary particles that really appear to have no finite extent. However, in week 4 we showed that it also applies to *systems* of particles, with the replacement of the position of the particle by the position of the center of mass of the system and the force with the total *external* force acting on the entire system (internal forces cancelled), and to *extended objects* made up of many of those infinitesimal chunklets. We could then extend Newton's First Law to apply as well to amorphous systems such as clouds of gas or structured systems such as "rigid objects" as long as we considered being "at rest" a statement concerning the motion of their *center of mass*. Thus a "baseball", made up of a truly staggering number of elementary microscopic particles, becomes a "particle" in its own right located at its center of mass.

We also learned that the *force* equilibrium of particles acted on by conservative force occurred at the points where the potential energy was maximum or minimum or neutral (flat), where we named maxima "unstable equilibrium points", minima "stable equilibrium points" and flat regions "neutral equilibria"[154].

However, in weeks 5 and 6 we learned enough to now be able to see that force equilibrium *alone* is ***not sufficient*** to cause an extended object or collection of particles to be in equilibrium. We can easily arrange situations where *two* forces act on an object in opposite directions (so there is no net force) but along lines such that together they exert a nonzero *torque* on the object and hence cause it to angularly accelerate and gain kinetic energy without bound, hardly a condition one would call "equilibrium".

The good news is that Newton's Second Law for Rotation is sufficient to imply Newton's First Law for Rotation:

> ***If, in an inertial reference frame, a rigid object is initially at rotational rest (not rotating), it will remain at rotational rest unless acted upon by a net external torque.***

That is, $\vec{\tau} = I\vec{\alpha} = 0$ implies $\vec{\Omega} = 0$ and constant[155]. We will call the condition where $\vec{\tau} = 0$ and a rigid object is not rotating ***torque equilibrium***.

We can make a baseball (initially with its center of mass at rest and not rotating) spin without exerting a net force on it that makes its center of mass move – it can be in force equilibrium but not torque equilibrium. Similarly, we can throw a baseball without imparting

[154] Recall that neutral equilibria were generally closer to being unstable than stable, as any nonzero velocity, no matter how small, would cause a particle to move continuously across the neutral region, making no particular point *stable* to say the least. That same particle would *oscillate*, due to the restoring force that traps it between the *turning points* of motion that we previously learned about and at least remain in the "vicinity" of a true stable equilibrium point for small enough velocities/kinetic energies.

[155] Whether or not I is a scalar or a tensor form...

any rotational spin – it can be in torque equilibrium but not force equilibrium. If we want the baseball (or any rigid object) to be in a *true* static equilibrium, one where it is *neither* translating *nor* rotating in the future if it is at rest and not rotating initially, we need both the conditions for force equilibrium and torque equilibrium to be true.

Therefore we now *define* the conditions for the ***static equilibrium of a rigid body*** to be:

> ***A rigid object is in static equilibrium when both the vector torque and the vector force acting on it are zero.***

That is:

> ***If $\vec{F}_{\rm tot} = 0$ and $\vec{\tau}_{\rm tot} = 0$, then an object initially at translational and rotational rest will remain at rest and neither accelerate nor rotate.***

That's it. Really, pretty simple.

Needless to say, the *idea* of stable versus unstable and neutral equilibrium still holds for torques as well as forces. We will consider an equilibrium to be stable only if ***both*** the force ***and*** the torque are "restoring" – and push or twist the system *back* to the equilibrium if we make small linear or angular displacements away from it.

7.2: Changing Frames and the Invariance of Equilibrium

Suppose you are looking at a rigid object that is in static equilibrium in an inertial reference frame where it is neither translating nor rotating and we compute all torquest from a pivot at its origin. Clearly, if we *change frames* to any frame that is at rest (not translating or rotating) relative to the first frame, we expect it to *still* be in equilibrium in the new frame.

However, the torque due to *each* force acting on the object *changes* as we move the frame origin – and pivot – around! For a single force acting at a point on the object, for example, the new torque might be zero (if the new pivot chosen lines up with the force or is located at the point where that force is applied) or *non*-zero if it doesn't! How can we be sure that equilibrium in one frame is equilibrium in all frames if the torque shifts with pivot?

Let's prove this! Start with static equilibrium in the first frame S:

$$\vec{\boldsymbol{F}}_1 + \vec{\boldsymbol{F}}_2 + \vec{\boldsymbol{F}}_3 ... = 0 \tag{7.1}$$

$$\vec{\boldsymbol{r}}_1 \times \vec{\boldsymbol{F}}_1 + \vec{\boldsymbol{r}}_2 \times \vec{\boldsymbol{F}}_2 + \vec{\boldsymbol{r}}_3 \times \vec{\boldsymbol{F}}_3 + ... = 0 \tag{7.2}$$

If we shift the origin/pivot by a constant vector $\vec{\boldsymbol{r}}_s$, to generate a new inertial reference frame S' (as we have done several times now) then the force sum obviously is still zero, but what happens to the total torque? All of the torques are measured with respect to the *new* origin of $S;$, so:

$$\vec{\boldsymbol{r}}_1{}' = \vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_s,\ \vec{\boldsymbol{r}}_2{}' = \vec{\boldsymbol{r}}_2 - \vec{\boldsymbol{r}}_s,\ \vec{\boldsymbol{r}}_3{}' = \vec{\boldsymbol{r}}_3 - \vec{\boldsymbol{r}}_s, ... \tag{7.3}$$

and in S':

$$\begin{aligned}
\vec{\boldsymbol{r}}_1{}' \times \vec{\boldsymbol{F}}_1 + \vec{\boldsymbol{r}}_2{}' \times \vec{\boldsymbol{F}}_2 + \vec{\boldsymbol{r}}_3{}' \times \vec{\boldsymbol{F}}_3 + ... &= \\
(\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_s) \times \vec{\boldsymbol{F}}_1 + (\vec{\boldsymbol{r}}_2 - \vec{\boldsymbol{r}}_s) \times \vec{\boldsymbol{F}}_2 + (\vec{\boldsymbol{r}}_3 - \vec{\boldsymbol{r}}_s) \times \vec{\boldsymbol{F}}_3 + ... &= \\
\cancel{\left\{ \vec{\boldsymbol{r}}_1 \times \vec{\boldsymbol{F}}_1 + \vec{\boldsymbol{r}}_2 \times \vec{\boldsymbol{F}}_2 + \vec{\boldsymbol{r}}_3 \times \vec{\boldsymbol{F}}_3 + ... \right\}} - \left\{ \vec{\boldsymbol{r}}_s \times \cancel{\left(\vec{\boldsymbol{F}}_1 + \vec{\boldsymbol{F}}_2 + \vec{\boldsymbol{F}}_3 ... \right)} \right\} &= 0 \quad (!)
\end{aligned}$$

(from equations 7.1 and 7.2.

This is a very simple proof that:

> ***Static equilibrium in one inertial reference frame is static equilibrium in all inertial reference frames!***

This is a *critical* little mini-theorem when it comes time to set up static equilibrium problems in the easiest possible way. In the examples below (and in your homework and other practice) you will often be required to choose an origin/pivot, and will rapidly learn to choose one that makes the problem *simpler* and solvable with *less algebra* (and hence fewer chances to make mistakes). The theorem above basically says that this is ok to do – while *the torques themselves* (unlike the forces) will change, possibly by quite a lot, as we change frames and pivots, if they sum to zero in one from they sum to zero in all frames, so you are free to choose a frame and pivot that makes the problem easy! Well, easi*er*, anyway...

7.3: Static Equilibrium Problems

After working so long and hard on solving actual *dynamical* problems involving force and torque, static equilibrium problems sound like they would be pretty trivial. After all, *nothing happens!* It seems as though solving for what happens when *nothing* happens would be easy.

Not so. To put it in perspective, let's consider *why* we might want to solve a problem in static equilibrium. Suppose we want to build a house. A properly built house is one that won't just *fall down*, either all by itself or the first time the wolf huffs and puffs at your door. It seems as though building a house that is *stable* enough not to fall down when you move around in it, or load it with furniture in different ways, or the first time a category 2 hurricane roars by overhead and whacks it with 160 kilometer per hour winds is a worthy design goal. You might even want it to survive earthquakes!

If you think that building a stable house is *easy*, I commend trying to build a house out of cards[156] . You will soon learn that balancing force loads, taking advantage of friction (or other "fastening" forces", avoiding unbalanced torques is all actually remarkably tricky, for all that we can learn to do it without solving actual equations. Engineers who want to build *serious* structures such as bridges, skyscrapers, radio towers, cars, airplanes, and so on spend a *lot* of time learning statics (and a certain amount of dynamics, because no structure in a dynamical world filled with Big Bad Wolves is truly "static") because it is *very expensive* when buildings, bridges, and so on fall down, when their structural integrity fails.

[156] Wikipedia: http://www.wikipedia.org/wiki/House of cards. Yes, even this has a wikipedia entry. Pretty cool, actually!

Stability is just as important for physicians to understand. The human body is not the world's most stable structure, as it turns out. If you have ever played with Barbie or G.I. Joe dolls, then you understand that getting them to stand up on their own takes a bit of doing – just a bit of bend at the waist, just a bit too much weight at the side, or the feet not adjusted just so, and they fall right over. Actual humans stabilize their erect stance by *constantly adjusting the force balance* of their feet, shifting weight without thinking to the heel or to the toes, from the left foot to the right foot as they move their arms or bend at the waist or lift something. Even healthy, coordinated adults who are paying attention nevertheless sometimes *lose their balance* because their motions exceed the fairly narrow tolerance for stability in some stance or another.

This ability to remain stable standing up or walking rapidly disappears as one's various sensory feedback mechanisms are impaired, and many, many health conditions impair them. Drugs or alcohol, neuropathy, disorders of the vestibular system, pain and weakness due to arthritis or aging. Many injuries (especially in the elderly) occur because people just plain *fall over*.

Then there is the fact that nearly all of our physical activity involves the adjustment of static balance between muscles (providing tension) and bones (providing compression), with our joints becoming stress-points that have to provide enormous forces, painlessly, on demand. In the end, physicians have to have a *very good conceptual understanding* of static equilibrium in order to help their patients acheive it and maintain it in the many aspects of the "mechanical" operation of the human body where it is essential.

For this reason, no matter who you are taking this course, you need to learn to solve real statics problems, ones that can help you later understand and work with statics as your life and career demand. This may be nothing more than helping your kid build a stable tree house, but y'know, you don't want to help them build a tree house and have that house *fall out of the tree* with your kid in it, nor do you want to deny them the joy of hanging out in their own tree house up in the trees!

As has been our habit from the beginning of the course, we start by considering the simplest problems in static equilibrium and then move on to more difficult ones. The simplest problems cannot, alas, be truly *one dimensional* because if the forces involved are truly one dimensional (and act to the right or left along a single line) there *is* no possible torque and force equilibrium suffices for both.

The simplest problem involving both force *and* torque is therefore at least three dimensional – two dimensional as far as the forces are concerned and one dimensional as far as torque and rotation is concerned. In other words, it will involve force balance in some plane of (possible) rotation and torque balance perpendicular to this plane alone a (possible) axis of rotation.

Example 7.3.1: Balancing a See-Saw

One typical problem in statics is balancing weights on a see-saw type arrangement – a uniform plank supported by a fulcrum in the middle. This particular problem is really only one dimensional as far as force is concerned, as there is no force acting in the x-direction or z-direction. It is one dimensional as far as torque is concerned, with rotation around any pivot one might

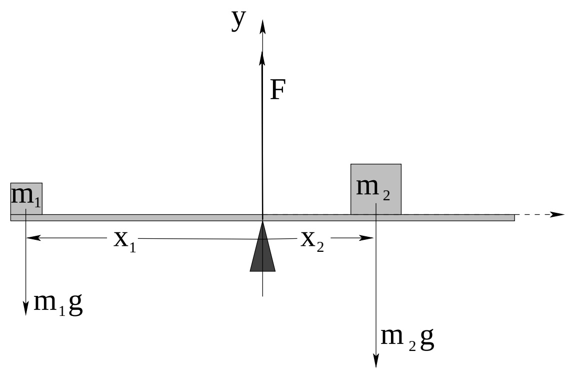


Figure 7.1: You are given m_1, x_1, and x_2 and are asked to find m_2 and F such that the see-saw is in *static equilibrium*.

select either into or out of the paper.

Static equilibrium requires force balance (one equation) and torque balance (one equation) and therefore we can solve for pretty much any two variables (unknowns) visible in figure 7.1 above. Let's imagine that in *this* particular problem, the mass m_1 and the distances x_1 and x_2 are given, and we need to find m_2 and F.

We have two choices to make – where we select the pivot[157] and which direction (in or out of the page) we are going to define to be "positive". A perfectly reasonable (but not unique or necessarily "best") choice is to select the pivot at the fulcrum of the see-saw where the unknown force F is exerted, and to select the $+z$-axis as positive rotation (out of the page as drawn).

We then write:

$$\sum F_y = F - m_1 g - m_2 g = 0 \tag{7.4}$$

$$\sum \tau_z = x_1 m_1 g - x_2 m_2 g = 0 \tag{7.5}$$

This is almost embarrassingly simple to solve. From the second equation:

$$m_2 = \frac{m_1 g x_1}{g x_2} = \left(\frac{x_1}{x_2}\right) m_1 \tag{7.6}$$

It is worth noting that this is precisely the mass that moves the *center of mass* of the system so that it is square over the fulcrum/pivot.

From the first equation and the solution for m_2:

$$F = m_1 g + m_2 g = m_1 g \left(1 + \left(\frac{x_1}{x_2}\right)\right) = m_1 g \left(\frac{x_1 + x_2}{x_2}\right) \tag{7.7}$$

That's all there is to it! Obviously, we could have been given m_1 and m_2 and x_1 and been asked to find x_2 and F, etc, just as easily.

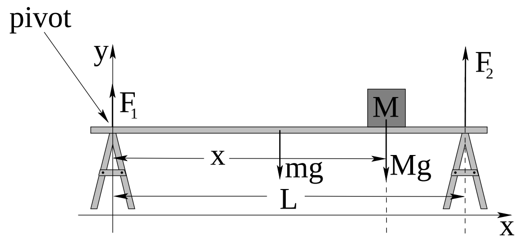


Figure 7.2: Two saw horses separated by a distance L support a plank of mass m symmetrically placed across them as shown. A block of mass M is placed on the plank a distance x from the saw horse on the left.

Example 7.3.2: Two Saw Horses

In figure 7.2, two saw horses separated by a distance L support a symmetrically placed plank. The rigid plank has mass m and supports a block of mass M placed a distance x from the left-hand saw horse. Find F_1 and F_2, the upward (normal) force exerted by each saw horse in order for this system to be in static equilibribum.

First let us pick something to put into equilibrium. The saw horses look pretty stable. The mass M does need to be in equilibrium, but that is pretty trivial – the plank exerts a normal force on M equal to its weight. The only tricky thing is the plank itself, which could and would rotate or collapse if F_1 and F_2 don't correctly balance the load created by the weight of the plank plus the weight of the mass M.

Again there are no forces in x or z, so we simply ignore those directions. In the y direction:

$$F_1 + F_2 - mg - Mg = 0 \tag{7.8}$$

or "the two saw horses must support the total weight of the plank plus the block", $F_1 + F_2 = (m + M)g$. This is not unreasonable or even unexpected, but it doesn't tell us how this weight is distributed between the two saw horses.

Once again we much choose a pivot. Four possible points – the point on the left-hand saw horse where F_1 is applied, the point at $L/2$ that is the center of mass of the plank (and half-way in between the two saw horses), the point under mass M where its gravitational force acts, and the point on the right-hand saw horse where F_2 is applied. Any of these will eliminate the torque due to one of the forces and presumably will simplify the problem relative to more arbitrary points. We select the left-hand point as shown – why not?

Then:

$$\tau_z = F_2 L - mgL/2 - Mgx = 0 \tag{7.9}$$

states that the torque around this pivot must be zero. We can easily solve for F_2:

$$F_2 = \frac{mgL/2 + Mgx}{L} = \frac{mg}{2} + Mg\frac{x}{L} \tag{7.10}$$

Finally, we can solve for F_1:

$$F_1 = (m + M)g - F_2 = \frac{mg}{2} + Mg\frac{L - x}{L} \tag{7.11}$$

[157]Which we now know is a *free choice* thanks to the theorem proved in the last section...

Does this make sense? Sure. The two saw horses *share* the weight of the symmetrically placed plank, obviously. The saw horse *closest* to the block M supports most of its weight, in a completely symmetric way. In the picture above, with $x > L/2$, that is saw horse 2, but if $x < L/2$, it would have been saw horse 1. In the middle, where $x = L/2$, the two saw horses symmetrically share the weight of the block as well!

This picture and solution are worth studying until all of this makes sense. Carrying things like sofas and tables (with the load shared between a person on either end) is a frequent experience, and from the solution to this problem you can see that if the load is not *symmetrical*, the person closest to the center of gravity will carry the largest load.

Let's do a slightly more difficult one, one involving equilibrium in *two* force directions (and one torque direction). This will allow us to solve for *three* unknown quantities.

Example 7.3.3: Hanging a Tavern Sign

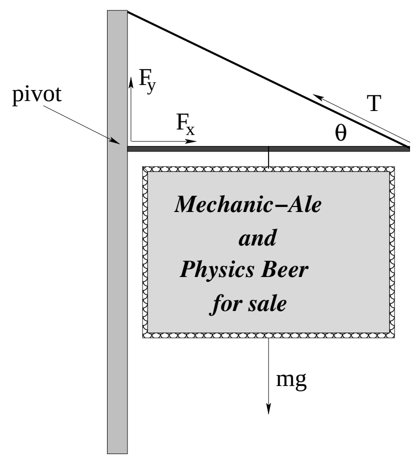


Figure 7.3: A sign with mass m is hung from a massless rigid pole of length L attached to a post and suspended by means of a wire at an angle θ relative to the horizontal.

Suppose that one day you get tired being a hardworking professional and decide to give it all up and open your own tavern/brewpub. Naturally, you site it in a lovely brick building close to campus (perhaps in Brightleaf Square). To attract passers-by you need a really good *sign* – the old fashioned sort made out of solid oak that hangs from a pole, one that (with the pole) masses $m = 50$ kg.

However, you *really* don't want the sign to either punch through the brick wall or break the suspension wire you are going to use to support the end farthest from the wall and you don't trust your architect because he seems way too interested in your future wares, so you decide to work out for yourself just what the forces are that the wall and wire have to support, as a function of the angle θ between the support pole and the support wire.

The physical arrangement you expect to end up with is shown in figure 7.3. You wish to find F_x, F_y and T, given m and θ.

By now the idea should be sinking in. Static equilibrium requires $\sum \vec{F} = 0$ and $\sum \vec{\tau} = 0$. There are no forces in the z direction so we ignore it. There is *only* torque in the $+z$ direction.

In this problem there is a clearly-best pivot to choose – one at the point of contact with the wall, where the two forces F_x and F_y are exerted. If we choose *this* as our pivot, these forces will not contribute to the net torque!

Thus:

$$\begin{aligned} F_x - T\cos(\theta) &= 0 && (7.12)\\ F_y + T\sin(\theta) - mg &= 0 && (7.13)\\ T\sin(\theta)L - mgL/2 &= 0 && (7.14) \end{aligned}$$

The last equation involves only T, so we solve it for:

$$T = \frac{mg}{2\sin(\theta)} \quad (7.15)$$

We can substitute this into the first equation and solve for F_x:

$$F_x = \frac{1}{2}mg\cot(\theta) \quad (7.16)$$

Ditto for the second equation:

$$F_y = mg - \frac{1}{2}mg = \frac{1}{2}mg \quad (7.17)$$

There are several features of interest in this solution. One is that the wire and the wall support each must support half of the weight of the sign. However, in order to accomplish this, the tension in the wire will be strictly greater than half the weight!

Consider $\theta = 30^\circ$. Then $T = mg$ (the entire weight of the sign) and $F_x = \frac{\sqrt{3}}{2}mg$. The magnitude of the force exerted by the wall on the pole equals the tension.

Consider $\theta = 10^\circ$. Now $T \approx 2.9mg$ (which still must equal the magnitude of the force exerted by the wall. Why?). The smaller the angle, the larger the tension (and force exerted by/on the wall). Make the angle *too* small, and your pole will punch right through the brick wall!

7.3.1: Equilibrium with a Vector Torque

So far we've only treated problems where all of the forces and moment arms live in a *single plane* (if not in a single direction). What if the moment arms themselves live in a plane? What if the forces exert torques in different directions?

Nothing changes a whole lot, actually. One simply has to set each *component* of the force and torque to zero *separately*. If anything, it may give us more equations to work with, and hence the ability to deal with more unknowns, at the cost of – naturally – some algebraic complexity.

Static equilibrium problems involving multiple torque directions are actually rather common. Every time you sit in a chair, every time food is placed on a table, the legs increase the forces they supply to the seat to maintain force *and* torque equilibrium. In fact, every time *any* two-dimensional sheet of mass, such as a floor, a roof, a tray, a table is suspended horizontally, one must solve a problem in vector torque to keep it from rotating around *any* of the axes in the plane.

We don't need to solve the most algebraically complex problems in the Universe in order to learn how to balance both multiple force components and multiple torque components, but we do need to do one or two to get the idea, because nearly everybody who is taking this course needs to be able to actually work with static equilibrium in multiple dimensions. Physicists need to be able to understand it both to teach it and to prime themselves for the full-blown theory of angular momentum in a more advanced course. Engineers, well, we don't want those roofs to tip *over*, those bridges to fall *down*. Physicians and veterinarians – balancing human or animal bodies so that they don't tip over one way *or another* seems like a good idea. Things like canes, four point walkers, special support shoes all are tools you may one day use to help patients retain the precious ability to navigate the world in an otherwise precarious vertical static equilibrium.

Example 7.3.4: Building a Deck

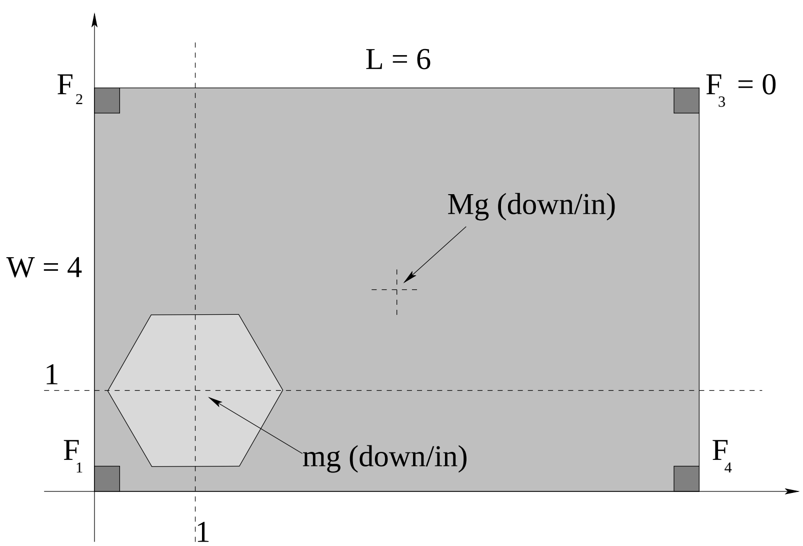


Figure 7.4

In figure 7.4 a very simple deck layout is shown. The deck is 4 meters wide and 6 meters long. It is supported by four load-bearing posts, one at each corner. You would like to put a hot tub on the deck, one that has a loaded mass of $m = 2000$ kg, so that its center of mass is 1 meter in and 1 meter up from the lower left corner (as drawn) next to the house. The deck itself is a uniform concrete slab of mass $M = 4000$ kg with its center of mass is at (3,2) from the lower left corner.

You would like to know if putting the hot tub on the deck will exceed the safe load capacity of the nearest corner support. It seems to you that as it is loaded with the hot tub, it will actually reduce the load on F_3. So find F_1, F_2, and F_4 when the deck is loaded in this way, assuming a perfectly rigid plane deck and $F_3 = 0$.

First of all, we need to choose a pivot, and I've chosen a fairly obvious one – the lower left corner, where the x axis runs through F_1 and F_4 and the y axis runs through F_1 and F_2.

Second, we need to note that $\sum F_x$ and $\sum F_y$ can be ignored – there are no lateral forces at all at work here. Gravity pulls the masses down, the corner beams push the deck itself up.

We can solve the normal force and force transfer in our heads – supporting the hot tub, the deck experiences a force equal to the *weight* of the hot tub right below the center of mass of the hot tub.

This gives us only *one* force equation:

$$F_1 + F_2 + F_4 - mg - Mg = 0 \tag{7.18}$$

That is, yes, the three pillars we've selected must support the total weight of the hot tub and deck together, since the F_3 pillar refuses to help out.

It gives us *two* torque equations, as hopefully it is obvious that $\tau_z = 0$! To make this nice and algebraic, we will set $h_x = 1$, $h_y = 1$ as the position of the hot tub, and use $L/2$ and $W/2$ as the position of the center of mass of the deck.

$$\sum \tau_x \;=\; WF_2 - \frac{W}{2}Mg - h_y mg \;=\; 0 \tag{7.19}$$

$$\sum \tau_y \;=\; h_x mg + \frac{L}{2}Mg - LF_4 \;=\; 0 \tag{7.20}$$

Note that if F_3 was acting, it would contribute to both of these torques and to the force above, and there would be an infinite number of possible solutions. As it is, though, solving this is pretty easy. Solve the last two equations for F_2 and F_4 respectively, then substitute the result into F_1. Only at the end substitute numbers in and see roughly what F_1 might be. Bear in mind that 1000 kg is a "metric ton" and weighs roughly 2200 pounds. So the deck and hot tub together, in this not-too-realistic problem, weigh over 6 tons!

Oops, we forgot the people in the hot tub and the barbecue grill at the far end and the furniture and the dog and the dancing. Better make the corner posts twice as strong as they need to be. Or even four times.

Once you see how this one goes, you should be ready to tackle the homework problem involving three legs, a tabletop, and a weight – same problem, really, but more complicated numbers.

7.4: Tipping

Another important application of the ideas of static equilibrium is to ***tipping problems***. A tipping problem is one where one uses the ideas of static equilibrium to identify the particular angle or force combination that will *marginally* cause some object to tip over. Sometimes this is presented in the context of objects on an inclined plane, held in place by static friction, and a tipping problem can be combined with a ***slipping problem***: determining if a block placed on an incline that is gradually raised tips first or slips first.

The idea of tipping is simple enough. An object placed on a flat surface is typically stable as long as the center of gravity is vertically ***inside the edges*** that are in contact with the surface, so that the torque created by the gravitational force around this limiting pivot is opposed by the torque exerted by the (variable) normal force.

That's all there is to it! Look at the center of gravity, look at the corner or edge – intuition will then tell you the object will "tip over", done.

Note that there *are* alternative kinds of tipping problems – ones where e.g. non-gravitational forces exerted from the *side* cause the object to tip over. Even here, however, one can generalize the intution gained from this example – if the total force swings past the tipping point so that normal forces can no longer oppose torquest about that point as the pivot, it will tip over.

Example 7.4.1: Tipping Versus Slipping

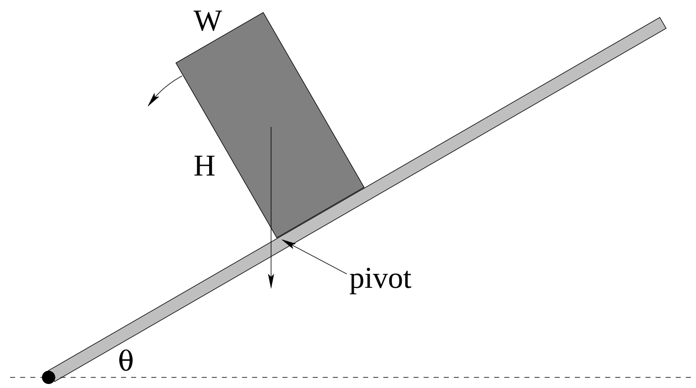


Figure 7.5: A rectangular block either tips first or slips (slides down the incline) first as the incline is gradually increased. Which one happens first? The figure is show with the block just past the ***tipping angle***.

In figure 7.5 a rectangular block of height H and width W is sitting on a rough plank that is gradually being raised at one end (so the angle it makes with the horizontal, θ, is slowly increasing).

At some angle we know that the block will start to slide. This will occur because the normal force is decreasing with the angle (and hence, so is the maximum force static friction can exert) and at the same time, the component of the weight of the object that points *down* the incline is increasing. Eventually the latter will exceed the former and the block will slide.

However, at some angle the block will *also* tip *over*. We know that this will happen because the normal force can only prevent the block from rotating *clockwise* (as drawn) around the pivot consisting of the lower left corner of the block. Unless the block has a magnetic lower surface, or a lower surface covered with velcro or glue, the plank cannot *attract* the lower surface of the block and prevent it from rotating *counterclockwise*.

As long as the net torque due to *gravity* (about this lower left pivot point) is *into* the page, the plank itself can exert a countertorque out of the page sufficient to keep the block from rotating down through the plank. If the torque due to gravity is *out of the page* – as it is in the figure 7.5 above, when the center of gravity moves over and to the *vertical left* of the pivot corner – the normal force exerted by the plank cannot oppose the counterclockwise torque of gravity and the block will fall over.

The ***tipping point***, or ***tipping angle*** is thus the angle where the ***center of gravity is directly over the pivot*** that the object will "tip" around as it falls over.

A very common sort of problem here is to determine whether some given block or shape will tip first or slip first. This is easy to find. First let's find the slipping angle θ_s. Let "down" mean "down the incline". Then:

$$\sum F_{\rm down} = mg\sin(\theta) - F_s = 0 \tag{7.21}$$

$$\sum F_{\perp} = N - mg\cos(\theta) = 0 \tag{7.22}$$

From the latter, as usual:

$$N = mg\cos(\theta) \tag{7.23}$$

and $F_s \le F_s^{\rm max} = \mu_s N$.

When

$$mg\sin(\theta_s) = F_s^{\rm max} = \mu_s mg\cos(\theta_s) \tag{7.24}$$

the force of gravity down the incline precisely balances the force of static friction. We can solve for the angle where this occurs:

$$\theta_s = \tan^{-1}(\mu_s) \tag{7.25}$$

Now let's determine the angle where it tips over. As noted, this is where the torque due to gravity around the pivot that the object will tip over changes sign from in the page (as drawn, stable) to out of the page (unstable, tipping over). This happens when the center of mass passes ***directly over the pivot***.

From inspection of the figure (which is drawn *very close* to the tipping point) it should be clear that the tipping angle θ_t is given by:

$$\theta_t = \tan^{-1}\left(\frac{W}{H}\right) \tag{7.26}$$

So, which one wins? The ***smaller*** of the two, θ_s or θ_t, of course – that's the one that happens *first* as the plank is raised. Indeed, since both are inverse tangents, the smaller of:

$$\mu_s, W/H \tag{7.27}$$

determines whether the system slips first or tips first, no need to actually *evaluate* any tangents or inverse tangents!

Example 7.4.2: Tipping While Pushing

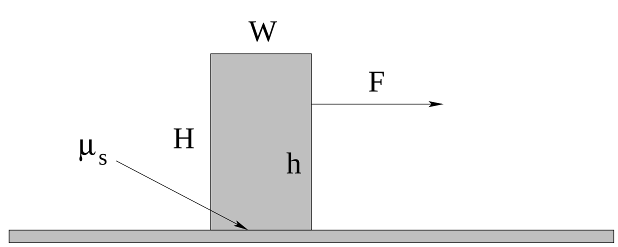


Figure 7.6: A uniform rectangular block with dimensions W by H (which has its center of mass at $W/2$, $H/2$) is pushed at a height h by a force F. The block sits on a horizontal smooth table with coefficient of static friction μ_s.

A uniform block of mass M being pushed by a horizontal outside force (say, a finger) a height h above the flat, smooth surface it is resting on (say, a table) as portrayed in figure 7.6.

If it is pushed down low (small h) the block slides. If pushed up high (large h) it tips. Find a condition for the height h at which it tips and slips at the same time.

The solution here is much the same as the solution to the previous problem. We *independently* determine the condition for slipping, as that is rather easy, and then using the maximum force that can be applied without it quite slipping, find the height h at which the block barely starts to tip over.

"Tipping over" in this case means that all of the normal force will be concentrated right at the pivot corner (the one the block will rotate around as it tips over) because the rest of the bottom surface is *barely* starting to leave the ground. All of the force of static friction is similarly concentrated at this one point. This is convenient to us, since neither one will therefore contribute to the torque around this point!

Conceptually, then, we seek the point h where, pushing with the maximum non-slipping force, the torque due to gravity *alone* is exactly equal to the torque exerted by the force F. This seems simple enough.

To find the force we need only to examine the usual force balance equations:

$$F - F_s = 0 \tag{7.28}$$
$$N - Mg = 0 \tag{7.29}$$

and hence $N = Mg$ (as usual) and:

$$F_{\max} = F_s^{\max} = \mu_s N = \mu_s Mg \tag{7.30}$$

(also as usual). Hopefully by now you had this completely solved in your head before I even wrote it all down neatly and were saying to yourself "$F_{\max}$, yeah, sure, that's $\mu_s Mg$, let's get on with it..."

So we shall. Consider the torque around the bottom right hand corner of the block (which is clearly and intuitively the "tipping pivot" around which the block will rotate as it falls over when the torque due to F is large enough to overcome the torque of gravity). Let us choose the positive direction for torque to be out of the page. It should then be quite obvious that when the block is *barely* tipping over, so that we can ignore any torque due to N and F_s:

$$\frac{WMg}{2} - h_{\rm crit} F_{\max} = \frac{WMg}{2} - h_{\rm crit}\mu_s Mg = 0 \tag{7.31}$$

or (solving for $h_{\rm crit}$, the critical height where it *barely* tips over even as it starts to slip):

$$h_{\rm crit} = \frac{W}{2\mu_s} \tag{7.32}$$

Now, does this make sense? If $\mu_s \to 0$ (a frictionless surface) we will never tip it *before* it starts to slide, although we might well push hard enough to tip it over in spite of it sliding. We note that in this limit, $h_{\rm crit} \to \infty$, which makes sense. On the other hand for finite μ_s if we let W become very small then $h_{\rm crit}$ similarly becomes very small, because the block is now very *thin* and is indeed rather precariously balanced.

The last bit of "sense" we need to worry about is $h_{\rm crit}$ compared to H. If $h_{\rm crit}$ is *larger* than H, this basically means that we *can't* tip the block over before it slips, for any reasonable μ_s.

This limit will always be realized for $W \gg H$. Suppose, for example, $\mu_s = 1$ (the upper limit of "normal" values of the coefficient of static friction that doesn't describe actual adhesion). Then $h_{\rm crit} = W/2$, and if $H < W/2$ there is no way to push in the block to make it tip before it slips. If μ_s is more reasonable, say $\mu_s = 0.5$, then only pushing at the very top of a block that is $W \times W$ in dimension marginally causes the block to tip. We can thus easily determine blocks "can" be tipped by a horizontal force and which ones cannot, just by knowing μ_s and looking at the blocks!

7.5: Force Couples

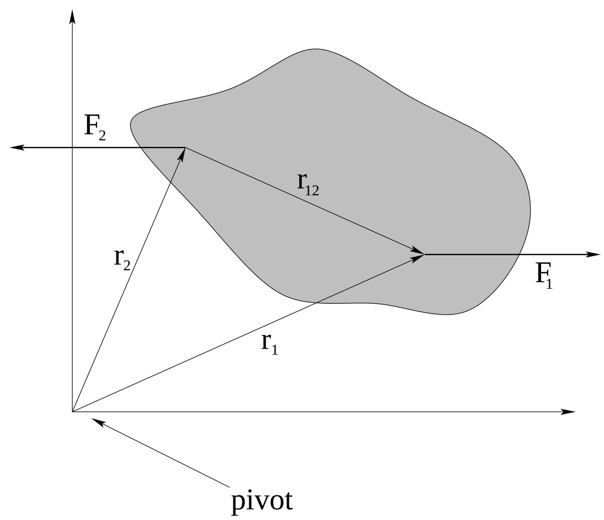


Figure 7.7: A ***Force Couple*** is a pair of equal and opposite forces that may or may not act along the line between the points where they are applied to a rigid object. Force couples exert a torque that is ***independent of the pivot*** on an object and (of course) do not accelerate the center of mass of the object.

Two equal forces that act in opposite directions but not necessarily along the same line are called a ***force couple***. Force couples are important both in torque and rotation problems and in static equilibrium problems. One doesn't have to be able to name them, of course – we know everything we need to be able to handle the physics of such a pair already without a name.

One important property of force couples does stand out as being worth deriving and learning on its own, though – hence this section. Consider the total torque exerted by a force couple in the coordinate frame portrayed in figure 7.7:

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{r}}_1 \times \vec{\boldsymbol{F}}_1 + \vec{\boldsymbol{r}}_2 \times \vec{\boldsymbol{F}}_2 \tag{7.33}$$

By hypothesis, $\vec{\boldsymbol{F}}_2 = -\vec{\boldsymbol{F}}_1$, so:

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{r}}_1 \times \vec{\boldsymbol{F}}_1 - \vec{\boldsymbol{r}}_2 \times \vec{\boldsymbol{F}}_1 = (\vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2) \times \vec{\boldsymbol{F}}_1 = \vec{\boldsymbol{r}}_{12} \times \vec{\boldsymbol{F}}_1 \tag{7.34}$$

This torque ***no longer depends on the coordinate frame!*** It depends only on the *difference* between $\vec{\boldsymbol{r}}_1$ and $\vec{\boldsymbol{r}}_2$, which is independent of coordinate system.

Note that we already used this property of couples when proving the law of conservation of angular momentum – it implies that internal Newton's Third Law forces can exert no torque on a system independent of intertial reference frame. Here it has a slightly different implication – it means that ***if the net torque produced by a force couple is zero in one frame, it is zero in all frames!*** The idea of static equilibrium itself is independent of frame!

It also means that equilibrium implies that ***the vector sum of all forces form force couples in each coordinate direction that are equal and opposite and that ultimately pass through the center of mass of the system***. This is a conceptually useful way to think about some tipping or slipping or static equilibrium problems.

Example 7.5.1: Rolling the Cylinder Over a Step

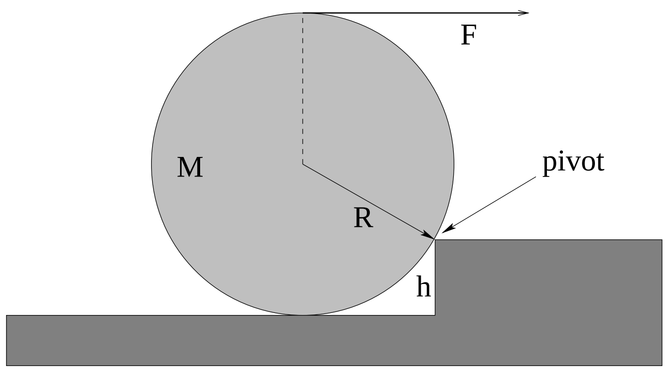


Figure 7.8

One classic example of static equilibrium and force couples is that of a ball or cylinder being rolled up over a step. The way the problem is typically phrased is:

a) Find the minimum force F that must be applied (as shown in figure 7.8) to cause the cylinder to *barely* lift up off of the bottom step and rotate up around the corner of the next one, assuming that the cylinder does not slip on the corner of the next step.

b) Find the force exerted by that corner at this marginal condition.

The simplest way to solve this is to recognize the point of the term "barely". When the force F is zero, gravity exerts a torque around the pivot *out* of the page, but the normal force of the tread of the lower stair exerts a countertorque precisely sufficient to keep the cylinder from rolling down into the stair itself. It also supports the weight of the cylinder. As F is increased, it exerts a torque around the pivot that is *into* the page, also opposing the gravitational torque, and the normal force *decreases* as less is needed to prevent rotation down into the step. At the same time, the pivot exerts a force that has to *both* oppose F (so the cylinder doesn't translate to the right) *and* support more and more of the weight of the cylinder as the normal force supports less.

At some particular point, the force exerted by the step *up* will precisely equal the weight of the step down. The force exerted by the step to the *left* will exactly equal the force F to the right. These forces will (vector) sum to zero and will incidentally exert no net torque either, as a pair of opposing couples.

That is enough that we could almost *guess* the answer (at least, if we drew some very nice pictures). However, we should work the problem algebraically to make sure that we all understand it. Let us assume that $F = F_m$, the desired minimum force where $N \to 0$. Then (with out of the page positive):

$$\sum \tau = mg\sqrt{R^2 - (R-h)^2} - F_m(2R - h) = 0 \tag{7.35}$$

where I have used the $r_\perp F$ form of the torque in both cases, and used the pythagorean theorem and/or inspection of the figure to determine $r_\perp$ for each of the two forces. No torque due to N is present, so F_m in this case is indeed the minimum force F at the marginal point where rotation *just* starts to happen:

$$F_m = \frac{mg\sqrt{R^2 - (R-h)^2}}{2R - h} \tag{7.36}$$

Next summing the forces in the x and y direction and solving for F_x and F_y exerted by the pivot corner itself we get:

$$F_x = -F_m = -\frac{mg\sqrt{R^2 - (R-h)^2}}{2R - h} \tag{7.37}$$

$$F_y = mg \tag{7.38}$$

Obviously, these forces form a perfect couple such that the torques and forces vanish!

To summarize, for problems like this look for the marginal static condition where rotation, or tipping, occurs. Set it up algebraically, and then solve! Finally, it is always a *good idea* to check dimensions and ask yourself if the answer is *reasonable*.

In other words, while the algebra above is enough to illustrate how to set up a problem like this when you encounter it in the in-class problem set and homework, but it fails to answer one very important question. We assumed – well, really the problem told us explicitly – that the cylinder did not slip on the step in order to obtain these answers, but is that really *reasonable*?

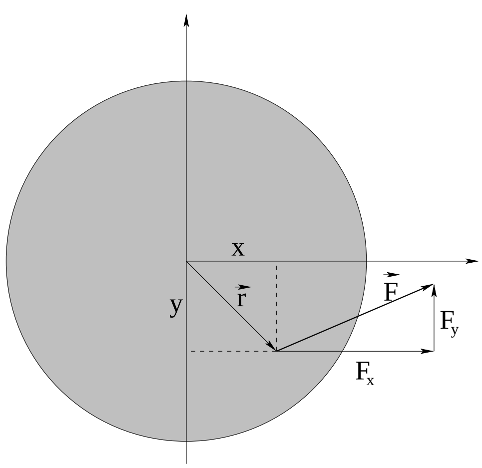
It turns out (see the *advanced* problem on the homework) that unless the step is quite shallow (that is, $h \ll R$) it is *not*. In the particular configuration shown (where we might estimate $h \approx R/2$) μ_s would have to be absurdly large for the cylinder to actually *roll* over the step corner without slipping on it instead before it actually lifts up off of the step! Showing this explicitly, however, is a bit of an algebraic challenge. The problems like this that you might actually be given on quizzes and exams will therefore be ones where this is not an issue,as I don't like giving unreasonable questions on exams!

Homework for Week 7

Problem 1.

Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

Problem 2.



a) In the figure to the left, a force $\vec{\boldsymbol{F}}$ is applied to a disk at point $\vec{\boldsymbol{r}}$ as shown, where:

$$\vec{\boldsymbol{F}} = 2\hat{\boldsymbol{x}} + 1\hat{\boldsymbol{y}} \text{ (newtons)}$$
$$\vec{\boldsymbol{r}} = 2\hat{\boldsymbol{x}} - 2\hat{\boldsymbol{y}} \text{ (meters)}$$

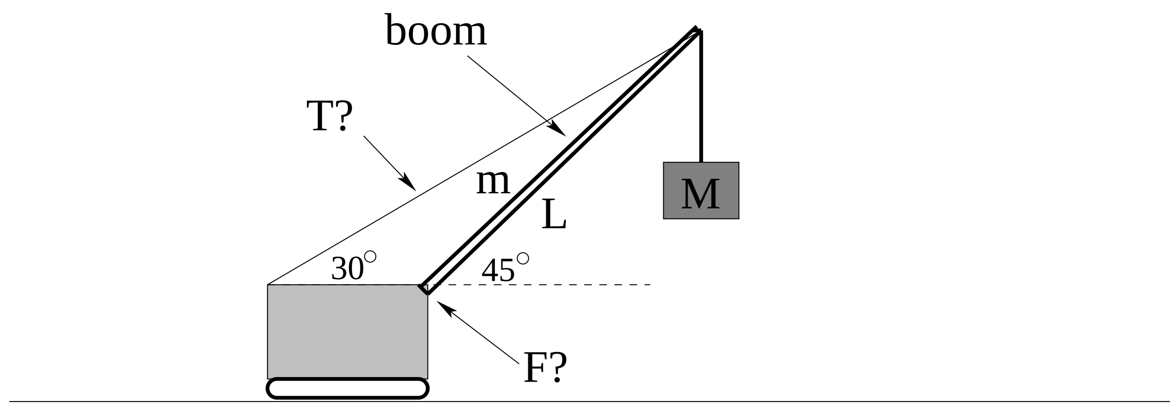
as shown. (That is, $F_x = 2$ N, $F_y = 1$ N, $x = 2$ m, $y = -2$ m).

Find the *total torque* about a pivot *at the origin*. Don't forget that torque is a ***vector***, so give the answer in cartesian components/vector form!

b) Find the ***cartesian*** vector torque for the following 3D force applied at the 3D position relative to a specified pivot (not shown):

$$\vec{\boldsymbol{F}} = 3\hat{\boldsymbol{x}} + 2\hat{\boldsymbol{y}} + 4\hat{\boldsymbol{z}} \text{ (newtons)}$$
$$\vec{\boldsymbol{r}} = -2\hat{\boldsymbol{x}} + 3\hat{\boldsymbol{y}} - 2\hat{\boldsymbol{z}} \text{ (meters)}$$

Problem 3.



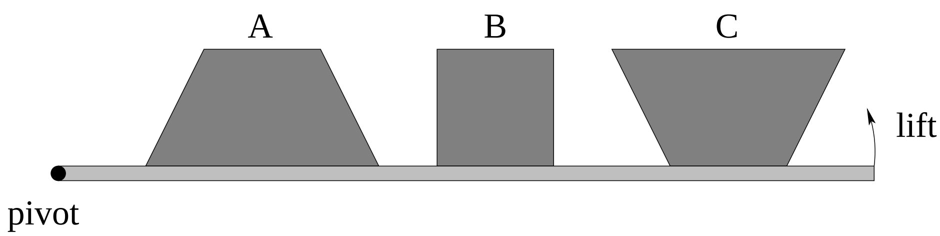
A crane with a boom (the long support between the body and the load) of mass m and length L holds a mass M suspended as shown. Assume that the center of mass of the boom is at $L/2$. Note that the wire with the tension T is **fixed** to the top of the boom, not run over a pulley to the mass M.

a) Find the magnitude of the tension T in the wire.

b) Find the ***vector*** force $\vec{F}$ exerted on the boom by the crane body.

You should *not need to use a calculator* if you use the cartesian form for the torque(s) and remember the sines and cosines for $45°$-$45°$-$90°$ and $30°$-$60°$-$90°$ triangles in the "obvious" choice for coordinate system and pivot!

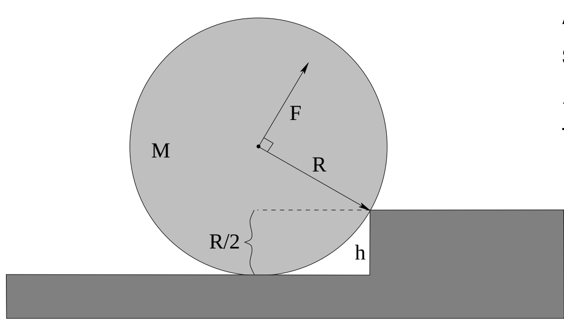
Problem 4.



In the figure above, three shapes (with uniform mass distribution and thickness, but possibly different masses) are drawn sitting on a plane that can be tipped up gradually. Assuming that static friction is great enough that all of these shapes will tip over before they slide, rank them ***in the order they will tip over as the angle of the board they are sitting on is increased***:

☐ before ☐ before ☐

Problem 5.



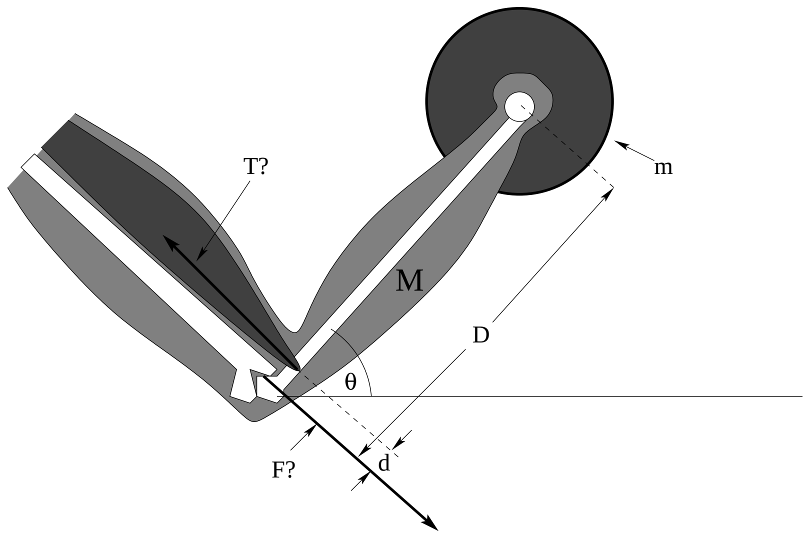
A cylinder of mass M and radius R sits against a step of height $\boldsymbol{h = R/2}$ as shown above. A force $\vec{F}$ is ***applied at right angles*** to the line connecting the corner of the step and the center of the cylinder.

a) Find the critical value $F_{\rm crit}$ such that any $F > F_{\rm crit}$ will roll the cylinder over the step if the cylinder does not slide on the corner. (Note: $F = |\vec{\boldsymbol{F}}|$, as usual.)

b) What is the vector force $\vec{\boldsymbol{F}}_c$ exerted by the corner (magnitude and direction) when the critical force found in part a) is applied to the disk?

All answers should be in terms of M, R, and g, as needed.

Problem 6.



An exercising human person holds their forearm of mass M and a barbell of mass m at rest at an angle θ with respect to the horizontal in an isometric curl as shown. Their bicep muscle that supports the suspended weight is connected ("inserted" in healthspeak) to the radius bone a short distance d up from the elbow joint and in the position shown, acts at right angles to it. The forearm bone[158] that supports the weight has length D.

a) Find (really ***estimate***) the tension T in the muscle, assuming for the moment that the center of mass of the forearm is in the middle at $D/2$.

b) Find the x and y components of the force exerted on the bone(s) of the forearm by the elbow joint in the geometry shown.

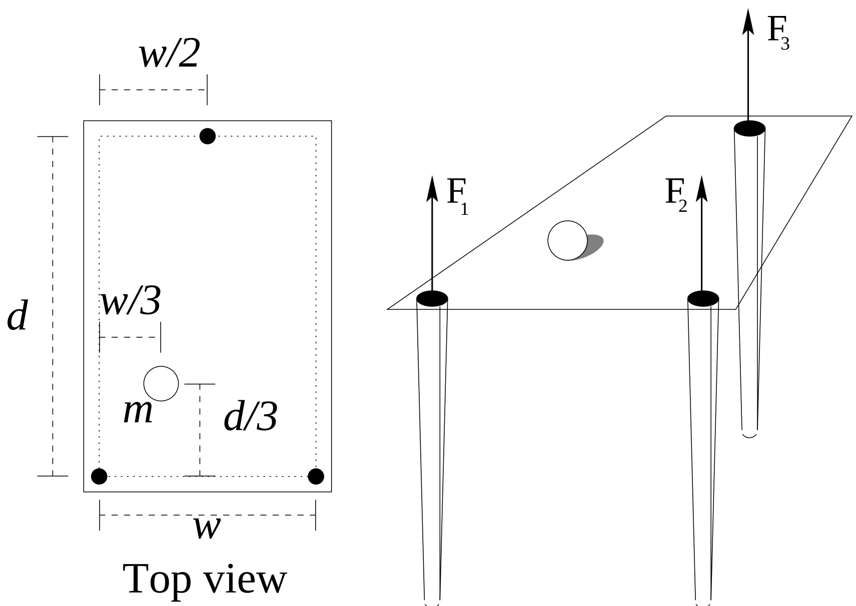
c) A fairly typical arm has $D \approx 30$ cm, $d \approx 3$ cm. Suppose that the mass of the forearm and weight are both 10 kg. Evaluate your answers for a) and b) when $\theta = \pi/4$.

Treat *both* components of this force as *unknowns* – do not assume that $\vec{F}$ points in the direction shown! It doesn't!

Also note that both T and the magnitude of F are ***much larger*** – by around an order of magnitude – than "just" the weight of the barbell/forearm being supported. This is one of several reasons joints wear out with age – they experience enormous stresses comparable to or greater than our entire weight concentrated at a point throughout the day even when our activity is rather mundane, over *years!*

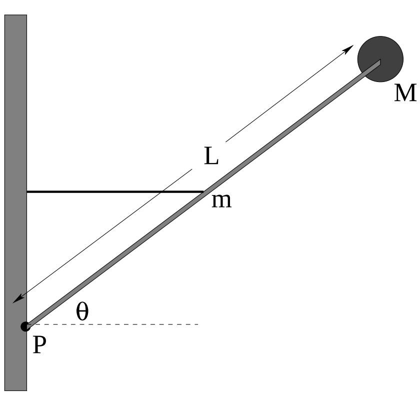
[158]The biceps muscle connects to the radius, one of the *two* bones in the forearm, but the radius and ulna (and the surrounding muscles) "move as one" for our estimate.

Problem 7.



The figure below shows a mass m placed on a table consisting of three narrow cylindrical legs at the positions shown with a light (presume massless) sheet of Plexiglas placed on top. First, determine what *kind* of problem it is (that is, what are the nontrivial force and torque equilibrium equations). Then use these equations to determine the magnitudes of the vertical forces F_i exerted by the Plexiglas on each leg (or vice versa) when the mass m s in the position shown?

Problem 8.

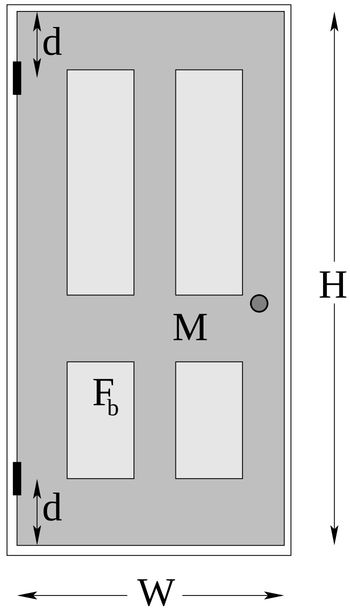


A small round mass M sits on the end of a rod of length L and mass m that is attached to a wall with a hinge at point P. The rod is kept from falling by a thin (massless) string attached horizontally between the midpoint of the rod and the wall. The rod makes an angle θ with the ground. Find:

a) the tension T in the string;

b) the force $\vec{F}$ exerted by the hinge on the rod.

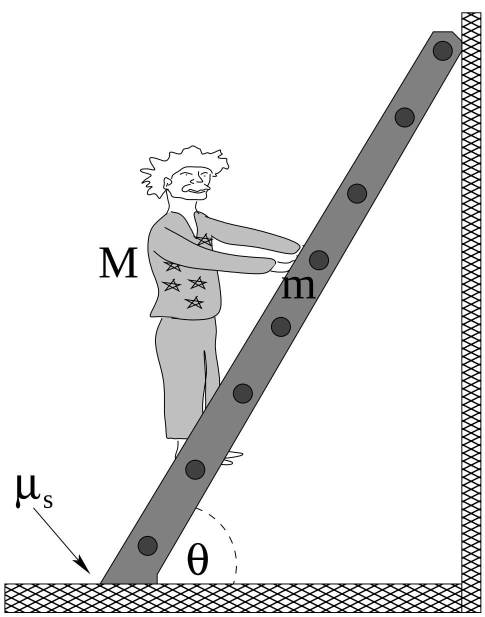
Problem 9.



A door of mass M that has height H and width W is hung from two hinges located a distance d from the top and bottom, respectively. ***Assuming that the vertical weight of the door is equally distributed between the two hinges***, find the total force (magnitude and direction) exerted by the top ($\vec{F}_t$) and bottom ($\vec{F}_b$) hinge.

Neglect the mass of the doorknob and assume that the center of mass of the door is at $W/2, H/2$. *Think* about how door hinges work before choosing directions for the hinge force components! Finally, *choose a good pivot!*

Problem 10.

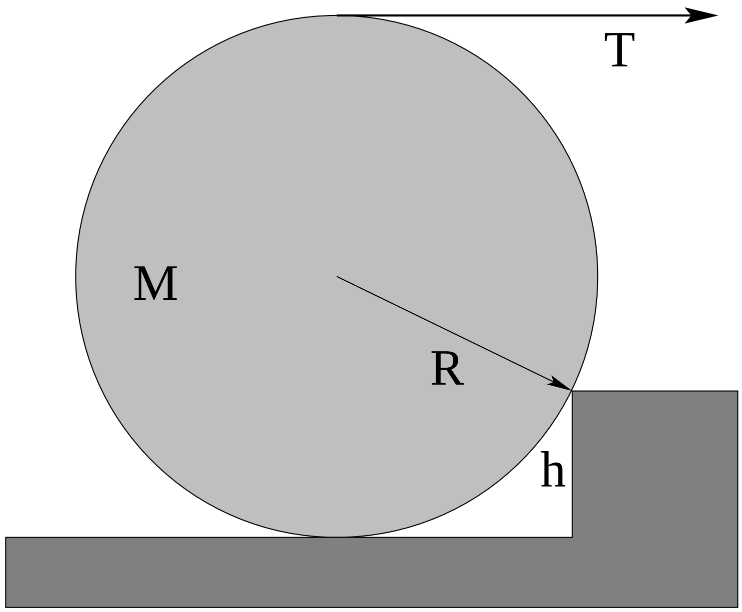


In the figure above, a ladder of mass m and length L is leaning against a wall at an angle θ. A person of mass M begins to climb the ladder. The ladder sits on the ground with a coefficient of static friction μ_s between the ground and the ladder. The wall is frictionless – it exerts only a normal force on the ladder.

Find the minimum angle $\theta_{\min}$ such that the person can climb the ladder all the way to the top without it slipping at the bottom.

*

Advanced Problem 11.



A cylinder of mass M and radius R sits against a step of height $h = R/2$ as shown above. A force $\vec{F}$ is applied parallel to the ground as shown. All answers should be in terms of M, R, g.

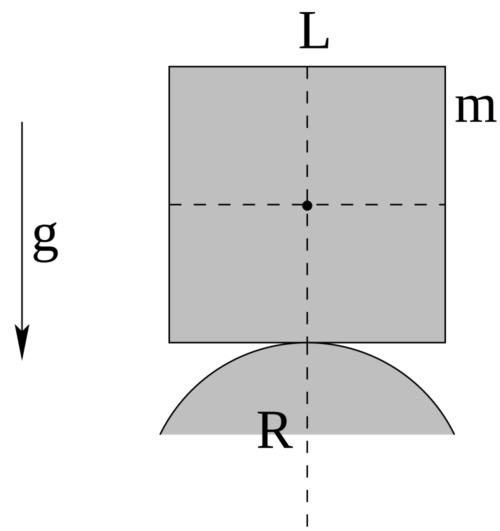
a) Find the minimum value of $|\vec{F}|$ that will roll the cylinder over the step if the cylinder does not slide on the corner.

b) What is the force exerted by the corner (magnitude and direction) when that force $\vec{F}$ is being exerted on the center?

c) Find the critical value $\mu_{s,\mathrm{crit}}$ such that if $\mu_s > \mu_{s,\mathrm{crit}}$, it will not slide on the corner!

*

Advanced Problem 12.



In the figure to the left, a cubic block of mass m is sitting on top of a round cylinder. It has a uniform density and center of mass located in its geometric center. The block is poised perfectly horizontally at the highest point on the cylinder and square to the page (so that the direction in and out of the page is irrelevant to this question). The radius of curvature of the cylinder is R, and the length of a side of the cube is L.

If $R \ll L$, this is like trying to balance the cube on a wire – the equilibrium is obviously *unstable* and it will tip over one way or the other unless it is *perfectly* balanced on top (which is non-physical, impossible in the real world). On the other hand, if $R \gg L$, the cube is *stable* and will always tip *back* to sit perfectly level on top of the cylinder if rolled a bit sideways around the curve.

Somewhere in between, there is a critical ratio such that the cube is stable if placed on top of the cylinder when:

$$\frac{L}{R} < \left(\frac{L}{R}\right)_{\rm crit}$$

That is, it will roll *back* to sit square at the top if rolled through a *small* angle around the cylinder *from* the top. **Find this critical ratio**.

III: Applications of Mechanics

Week 8: Fluids

1.14: Fluids Summary

- **Fluids** are states of matter characterized by a lack of long range order. They are characterized by their **density** ρ and their **compressibility**. Liquids such as water are (typically) relatively incompressible; gases can be significantly compressed. Fluids have other characteristics, for example viscosity (how strongly a fluid communicates *shear* stress from a boundary into its bulk volume) and adhesion (how "wet" water is, that is, how strongly it binds to material that is in contact with it at a confining surface).

- **Pressure** is the force per unit area exerted by a fluid on its surroundings:

$$P = F/A \tag{8.1}$$

  Its SI units are *pascals* where 1 pascal = 1 newton/meter squared. Pressure is also measured in "atmospheres" (the pressure of air at or near sea level) where 1 atmosphere $\approx 10^5$ pascals. The pressure in an incompressible fluid varies with depth according to:

$$P = P_0 + \rho g D \tag{8.2}$$

  where P_0 is the pressure at the top and D is the depth.

- **Pascal's Principle** Pressure applied to a fluid is transmitted undiminished to all points of the fluid.

- **Archimedes' Principle** The buoyant force on an object

$$F_b = \rho g V_{\rm disp} \tag{8.3}$$

  where frequency $V_{\rm disp}$ is the volume of fluid displaced by an object.

- **Conservation of Flow** We will study only steady/laminar flow in the absence of turbulence and viscosity.

$$I = A_1 v_1 = A_2 v_2 \tag{8.4}$$

  where I is the ***flow***, the volume per unit time that passes a given point in e.g. a pipe.

- For a circular smooth round pipe of length L and radius r carrying a fluid in ***laminar flow*** with ***dynamical viscosity***[159] μ, the flow is related to the pressure difference across the pipe by the ***resistance*** R:

$$\Delta P = IR \tag{8.5}$$

It is worth noting that this is the fluid-flow version of ***Ohm's Law***, which you will learn next semester if you continue. We will generally omit the modifier "dynamical" from the term viscosity in this course, although there is actually another, equivalent measure of viscosity called the ***kinematic viscosity***, $\nu = \mu/\rho$. The primary difference is the units – μ has the SI units of pascal-seconds where ν has units of meters square per second.

- The resistance R is given by the follow formula:

$$R = \frac{8L\mu}{\pi r^4} \tag{8.6}$$

and the flow equation above becomes ***Poiseuille's Law***[160] :

$$I = \frac{\Delta P}{R} = \frac{\pi r^4 \, \Delta P}{8L\mu} \tag{8.7}$$

The key facts from this series of definitions are that flow increases linearly with pressure (so to achieve a given e.g. perfusion in a system of capillaries one requires a sufficient pressure difference across them), increases with the ***fourth power*** of the radius of the pipe (which is why narrowing blood vessels become so dangerous past a certain point) and decreases with the length (longer blood vessels have a greater resistance).

- If we ***neglect resistance*** (an idealization roughly equivalent to neglecting friction) and consider the flow of fluid in a closed pipe that can e.g. go up and down, the ***work-mechanical energy theorem*** per unit volume of the fluid can be written as **Bernoulli's Equation**:

$$P + \frac{1}{2}\rho v^2 + \rho g h = \text{constant} \tag{8.8}$$

- **Venturi Effect** At constant height, the pressure in a fluid *decreases* as the velocity of the fluid *increases* (the work done by the pressure difference is what speeds up the fluid!. This is responsible for e.g. the lift of an airplane wing and the force that makes a spinning baseball or golf ball curve.

- **Torricelli's Rule:** If a fluid is flowing through a very small hole (for example at the bottom of a large tank) then the velocity of the fluid at the large end can be neglected in

[159] Wikipedia: http://www.wikipedia.org/wiki/viscosity. We will defer any actual statement of how viscosity is related to forces until we cover shear stress in a couple of weeks. It's just too much for now. Oh, and sorry about the symbol. Yes, we already have used μ for e.g. static and kinetic friction. Alas, we will use μ for still more things later. Even with both greek and roman characters to draw on, there just aren't enough characters to cover all of the quantities we want to algebraically work with, so you have to get used to their reuse ***in different contexts*** that hopefully make them easy enough to keep straight. I decided that it is better to use the accepted symbol in this textbook rather than make one up myself or steal a character from, say, Urdu or a rune from Ancient Norse.

[160] Wikipedia: http://www.wikipedia.org/wiki/Hagen-Poiseuille equation. The derivation of this result isn't *horribly* difficult or hard to understand, but it is long and beyond the scope of this course. Physics and math majors are encouraged to give it a peek though, if only to learn where it comes from.

Bernoulli's Equation. In that case the exit speed is the same as the speed of a mass dropped the same distance:

$$v = \sqrt{2gH} \tag{8.9}$$

where H is the depth of the hole relative to the top surface of the fluid in the tank.

8.1: General Fluid Properties

Fluids are the generic name given to two states of matter, liquids and gases[161] characterized by a lack of long range order and a high degree of mobility at the molecular scale. Let us begin by visualizing fluids microscopically, since we like to build our understanding of matter from the ground up.

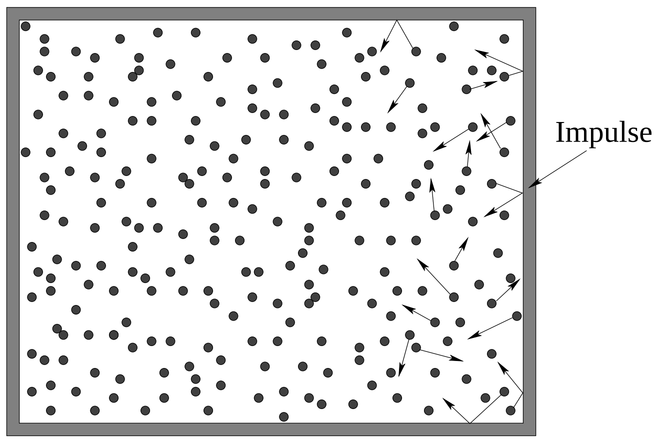


Figure 8.1: A large number of atoms or molecules are confined within in a "box", where they bounce around off of each other and the walls. They exert a *force* on the walls equal and opposite the the *force* the walls exert on them as the collisions more or less elastically reverse the particles' momenta perpendicular to the walls.

In figure 8.1 we see a highly idealized picture of what we might see looking into a tiny box full of gas. Many particles all of mass m are *constantly moving* in random, constantly changing directions (as the particles collide with each other and the walls) with an average kinetic energy related to the temperature of the fluid. Some of the particles (which might be atoms such as helium or neon or molecules such as H_2 or O_2) happen to be close to the walls of the container and moving in the right direction to bounce (elastically) off of those walls.

When they do, their *momentum* perpendicular to those walls is *reversed*. Since *many, many* of these collisions occur each second, there is a nearly continuous momentum transfer between the walls and the gas and the gas and the walls. This transfer, per unit time, becomes the *average force* exerted by the walls on the gas and the gas on the walls (see the problem in Week 4 with beads bouncing off of a pan).

[161]We will not concern ourselves with "plasma" as a possible fourth state of matter in this class, viewing it as just an "ionized gas" although a very dense plasma might well be more like a liquid. Only physics majors and perhaps a few engineers are likely to study plasmas, and you have plenty of time to figure them out after you have learned some electromagnetic theory.

Eventually, we will transform this simple picture into the ***Kinetic Theory of Gases*** and use it to derive the venerable ***Ideal Gas Law*** (physicist style)[162] :

$$PV = Nk_bT \tag{8.10}$$

but for now we will ignore the role of temperature and focus more on understanding the physical characteristics of the fluid such as its ***density***, the idea of ***pressure itself*** and the ***force*** exerted by fluids on themselves (internally) and on anything the fluid presses upon along the lines of the particles above and the walls of the box.

8.1.1: Pressure

As noted above, the walls of the container exert an average force on the fluid molecules that confine them by reversing their perpendicular momenta in collisions. The total momentum transfer is proportional to the number of molecules that collide per unit time, and this in turn is (all things being equal) clearly proportional to the area of the walls. Twice the surface area – when confining the same number of molecules over each part – has to exert twice the force as twice the number of collisions occur per unit of time, each transferring (on average) the same impulse. It thus makes sense, when considering fluids, to describe the forces that confine and act on the fluids in terms of ***pressure***, defined to be the ***force per unit area*** with which a fluid pushes on a confining wall or the confining wall pushes on the fluid:

$$P = \frac{F}{A} \tag{8.11}$$

Pressure gets its own SI units, which clearly must be Newtons per square meter. We give these units their own name, ***Pascals***:

$$1 \text{ Pascal} = \frac{\text{Newton}}{\text{meter}^2} \tag{8.12}$$

A Pascal is a tiny unit of pressure – a Newton isn't very big, recall (one kilogram weighs roughly ten Newtons or 2.2 pounds) so a Pascal is the weight of a quarter pound spread out over a square meter. Writing out "pascal" is a bit cumbersome and you'll see it sometimes abbreviated ***Pa*** (with the usual power-of-ten modifications, kPa, MPa, GPa, mPa and so on).

A more convenient measure of pressure in our everyday world is a form of the unit called a ***bar***:

$$1 \text{ bar} = 10^5 \text{ Pa} = 100 \text{ kPa} \tag{8.13}$$

As it happens, the average air pressure at sea level is *very nearly* 1 bar, and varies by at most a few percent on either side of this. For that reason, air pressure in the modern world is generally reported on the scale of ***millibars***, for example you might see air pressure given as 959 mbar (characteristic of the low pressure in a major storm such as a hurricane), 1023 mbar (on a fine, sunny day).

The mbar is probably the "best" of these units for describing everyday air pressure (and its temporal and local and height variation without the need for a decimal or power of ten), with Pascals being an equally good and useful general purpose arbitrary precision unit

[162] Wikipedia: http://www.wikipedia.org/wiki/Ideal Gas Law.

The symbol ***atm*** stands for ***one standard atmosphere***. The connection between atmospheres, bars, and pascals is:

$$1 \text{ standard atmosphere} = 101.325 \text{ kPa} = 1013.25 \text{ mbar} \tag{8.14}$$

Note that *real* air pressure at sea level is most unlikely to be this exact value, and although this pressure is often referred to in textbooks and encylopedias as "the average air pressure at sea level" this is not, in fact, the case. The extra significant digits therefore refer only to a fairly arbitrary value (in pascals) historically related to the original definition of a standard atmosphere in terms of "millimeters of mercury" or *torr*:

$$1 \text{ standard atmosphere} = 760.00 \text{ mmHg} = 760.00 \text{ torr} \tag{8.15}$$

that is of no practical or immediate use. All of this is discussed in some detail in the section on barometers below.

In this class we will use the simple rule 1 bar $\approx$ 1 atm to avoid having to divide out the extra digits, just as we approximated $g \approx 10$ when it is really closer to 9.8. This rule is more than adequate for nearly all purposes and makes pressure arithmetic something you can often do with fingers and toes or the back of an envelope, with around a 1% error if somebody actually gave a pressure in atmospheres with lots of significant digits instead of the superior pascal or bar SI units.

Note well: in the field of medicine ***blood pressures*** are given in mm of mercury (or torr) by long standing tradition (largely because for at least a century blood pressure was measured with a mercury-based sphygmomanometer). This is discussed further in the section below on the human heart and circulatory system. These can be converted into atmospheres by dividing by 760, remembering that one is measuring the *difference* between these pressures and the standard atmosphere (so the actual blood pressure is always *greater* than one atmosphere).

Pressure isn't only exerted at the *boundaries* of fluids. Pressure also describes the internal transmission of forces *within* a fluid. For example, we will soon ask ourselves "Why don't fluid molecules all fall to the ground under the influence of gravity and stay there?" The answer is that (at a sufficient temperature) the *internal pressure* of the fluid suffices to *support* the fluid above upon the back (so to speak) of the fluid below, all the way down to the ground, which of course has to support the weight of the entire column of fluid. Just as "tension" exists in a stretched string at all points along the string from end to end, so the pressure within a fluid is well-defined at all points from one side of a volume of the fluid to the other, although in neither case will the tension or pressure in general be *constant*.

8.1.2: Density

As we have done from almost the beginning, let us note that even a very tiny volume of fluid has many, many atoms or molecules in it, at least under ordinary circumstances in our everyday lives. True, we can work to create a *vacuum* – a volume that has relatively *few* molecules in it per unit volume, but it is almost impossible to make that number zero – even the hard vacuum of outer space has on average one molecule per cubic meter or thereabouts[163] . We live at

[163] Wikipedia: http://www.wikipedia.org/wiki/Vacuum. Vacuum is, of course, "nothing", and if you take the time to read this Wikipedia article on it you will realize that even *nothing* can be pretty amazing. In man-made vacuums, there are nearly always as many as hundreds of molecules per cubic centimeter.

the bottom of a gravity well that confines our *atmosphere* – the air that we breathe – so that it forms a relatively thick soup that we move through and breathe with order of *Avogadro's Number* (6×10^{23}) molecules per *liter* – hundreds of billions of billions per cubic centimeter.

At this point we cannot possibly track the motion and interactions of all of the individual molecules, so we *coarse grain* and *average*. The coarse graining means that we once again consider volumes that are large relative to the sizes of the atoms but small relative to our macroscopic length scale of meters – cubic millimeters or cubic microns, for example – that are large enough to contain many, many molecules (and hence a well defined *average* number of molecules) but small enough to treat like a differential volume for the purposes of using calculus to add things up.

We could just count molecules in these tiny volumes, but the properties of *oxygen* molecules and *helium* molecules might well be very different, so the molecular count alone may not be the most useful quantity. Since we are interested in how forces might act on these small volumes, we need to know their mass, and thus we define the *density* of a fluid to be:

$$\rho = \frac{dm}{dV}, \tag{8.16}$$

the ***mass per unit volume*** we are all familiar with from our discussions of the center of mass of continuous objects and moments of inertia of rigid objects.

Although the definition itself is the same, the density of a fluid behaves in a manner that is similar, but not quite identical in its properties, to the density of a solid. The density of a fluid usually varies *smoothly* from one location to another, because an excess of density in one place will spread out as the molecules travel and collide to smooth out, on average. The particles in some fluids (or almost any fluid at certain temperatures) are "sticky", or ***strongly interacting***, and hence the fluid coheres together in clumps where the particles are mostly touching, forming a *liquid*. In other fluids (or all fluids at higher temperatures) the molecules move so fast that they do *not* interact much and spend most of their time relatively far apart, forming a *gas*.

A gas spreads itself out to fill any volume it is placed in, subject only to forces that confine it such as the walls of containers or gravity. It assumes the shape of containers, and forms a (usually nearly spherical) layer of atmosphere around planets or stars when confined by gravity. Liquids also spread themselves out to some extent to fill containers they are placed in or volumes they are confined to by a mix of surface forces and gravity, but they also have the property of *surface tension* that can permit a liquid to exert a force of confinement on *itself*. Hence water fills a glass, but water also forms nearly spherical droplets when falling freely as surface tension causes the droplet to minimize its surface area relative to its volume, forming a sphere.

Surface chemistry or surface adhesion can also exert forces on fluids and initiate things like *capillary flow* of e.g. water up into very fine tubes, drawn there by the surface interaction of the hydrophilic walls of the tube with the water. Similarly, hydrophobic materials can actually repel water and cause water to bead up instead of spreading out to wet the surface. We will largely ignore these phenomena in this course, but they are very interesting and are actually *useful* to physicians as they use pipettes to collect fluid samples that draw themselves up into sample tubes as if by magic. It's not magic, it's just physics.

8.1.3: Compressibility

A major difference between fluids and solids, and liquids and gases within the fluids, is the *compressibility* of these materials. Compressibility describes how a material responds to changes in *pressure*. Intuitively, we expect that if we change the volume of the container (making it smaller, for example, by pushing a piston into a confining cylinder) while holding the amount of material inside the volume constant we will change the pressure; a smaller volume makes for a larger pressure. Although we are not quite prepared to derive and fully justify it, it seems at least *reasonable* that this can be expressed as a simple *linear relationship*:

$$\Delta P = -B\frac{\Delta V}{V} \tag{8.17}$$

Pressure up, volume down and vice versa, where the amount it goes up or down is related, not unreasonably, to the total volume that was present in the first place. The constant of proportionality B is called the ***bulk modulus*** of the material, and it is very much like (and closely related to) the *spring constant* in Hooke's Law for springs.

Note well that we haven't really specified *yet* whether the "material" is solid, liquid or gas. All three of them have densities, all three of them have bulk moduli. Where they differ is in the *qualitative* properties of their compressibility.

Solids are typically *relatively* incompressible (large B), although there are certainly exceptions. The have long range order – all of the molecules are packed and tightly bonded together in structures and there is usually very little free volume. Atoms themselves violently oppose being "squeezed together" because of the ***Pauli exclusion principle*** that forbids electrons from having the same set of quantum numbers as well as straight up ***Coulomb repulsion*** that you will learn about next semester.

Liquids are also relatively incompressible (large B). They differ from solids in that they lack long range order. All of the molecules are constantly moving around and any small "structures" that appear due to local interaction are short-lived. The molecules of a liquid are close enough together that there is often significant physical and chemical interaction, giving rise to surface tension and wetting properties – especially in water, which is (as one sack of water speaking to another) an amazing fluid!

Gases are in contrast quite *compressible* (small B). One can usually squeeze gases smoothly into smaller and smaller volumes, until they reach the point where the molecules are basically all touching and the gas converts to a liquid! Gases per se (especially hot gases) usually remain "weakly interacting" right up to where they become a liquid, although the correct (non-ideal) equation of state for a real gas often displays features that are the results of moderate interaction, depending on the pressure and temperature.

Water[164] is, as noted, a remarkable liquid. H_2O is a polar molecules with a permanent dipole moment, so water molecules are very strongly interacting, both with each other and with other materials. It organizes itself quickly into a state of relative order that is ***very incompressible***. The bulk modulus of water is 2.2×10^9 Pa, which means that even deep in the ocean where pressures can be measured in the tens of millions of Pascals (or hundreds of

[164] Wikipedia: http://www.wikipedia.org/wiki/Properties of Water. As I said, water is amazing. This article is well worth reading just for fun.

atmospheres) the density of water only varies by a few percent from that on the surface. Its density varies much more rapidly with *temperature* than with pressure[165]. We will idealize water by considering it to be *perfectly incompressible* in this course, which is close enough to true for nearly any mundane application of hydraulics that you are most unlikely to ever observe an exception that matters.

8.1.4: Viscosity and fluid flow

Fluids, whether liquid or gas, have some internal "stickiness" that resists the relative motion of one part of the fluid compared to another, a kind of internal "friction" that tries to equilibrate an entire body of fluid to move together. They also interact with the walls of any container in which they are confined. The ***viscosity*** of a fluid (symbol μ) is a measure of this internal friction or stickiness. Thin fluids have a low viscosity and flow easily with minimum resistance; thick sticky fluids have a high viscosity and resist flow.

Fluid, when flowing through (say) a cylindrical pipe tends to organize itself in one of two very different ways – a state of *laminar flow* where the fluid at the very edge of the flowing volume is at rest where it is in contact with the pipe and the speed concentrically and symmetrically increases to a maximum in the center of the pipe, and *turbulent flow* where the fluid tumbles and rolls and forms eddies as it flows through the pipe. Turbulence and flow and viscosity are properties that will be discussed in more detail below.

8.1.5: Properties Summary

To summarize, fluids have the following properties that you should conceptually and intuitively understand and be able to use in working fluid problems:

- They usually assume the shape of any vessel they are placed in (exceptions are associated with confinement due to gravity and surface effects such as surface tension and how well the fluid adheres to the surface in question).
- They are characterized by a mass per unit volume *density* ρ.
- They exert a *pressure* P (force per unit area) on themselves and any surfaces they are in contact with.
- The pressure can vary according to the dynamic and static properties of the fluid.
- The fluid has a measure of its "stickiness" and resistance to flow called *viscosity*. Viscosity is the internal friction of a fluid, more or less. We will treat fluids as being "ideal" and ignore viscosity in this course.

[165]A fact that impacts my beer-making activities quite significantly, as the specific gravity of hot wort fresh off of the boil is quite different from the specific gravity of the same wort cooled to room temperature. The specific gravity of the wort is related to the sugar content, which is ultimately related to the alcohol content of the fermented beer. Just in case this interests you...

- Fluids are *compressible* – when the pressure in a fluid is increased, its volume descreases according to the relation:

$$\Delta P = -B\frac{\Delta V}{V} \tag{8.18}$$

where B is called the ***bulk modulus*** of the fluid (the equivalent of a spring constant). An alternative formulation of this equation uses the ***compressibility*** β:

$$\beta = -\frac{1}{V}\left(\frac{dV}{dP}\right) = \frac{1}{B}$$

- Fluids where β is a very small number (so large changes in pressure create only tiny changes in fractional volume) are called ***incompressible*** and their density is roughly constant. Water is an example of an incompressible fluid, as are most liquids. Gases have a higher compressibility and pressure changes can produce a large change in occupied volume (and hence density).

- Below a critical speed, the dynamic flow of a moving fluid tends to be laminar, where every bit of fluid moves parallel to its neighbors in response to pressure differentials and around obstacles. Above that speed it becomes turbulent flow. Turbulent flow is quite difficult to treat mathematically and is hence beyond the scope of this introductory course – we will restrict our attention to ideal fluids either static or in laminar flow.

We will now use these general properties and definitions, plus our existing knowledge of physics, to deduce a number of important properties of and laws pertaining to *static fluids*, fluids that are in static equilibrium.

1.15: Static Fluids

8.1.6: Pressure and Confinement of Static Fluids

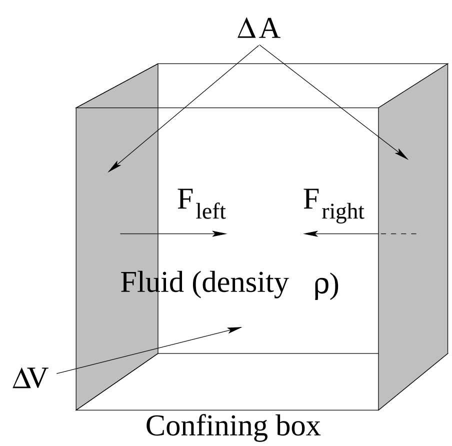


Figure 8.2: A fluid in static equilibrium confined to a sealed rectilinear box in zero gravity.

In figure 8.2 we see a box of a fluid that is confined within the box by the rigid walls of the box. We will imagine that this particular box is in “free space” far from any gravitational attractor and is therefore at rest with no external forces acting on it. We know from our intuition based on things like cups of coffee that no matter how this fluid is initially stirred up and moving *within* the container, after a very long time the fluid will damp down any initial motion by interacting with the walls of the container and arrive at *static equilibrium*[166].

A fluid in static equilibrium has the property that every single tiny chunk of volume in the fluid has to *independently* be in *force equilibrium* – the total force acting on the differential volume chunk must be *zero*. In addition the net torques acting on all of these differential subvolumes must be zero, and the fluid must be at rest, neither translating nor rotating. Fluid rotation is more complex than the rotation of a static object because a fluid can be *internally* rotating even if all of the fluid in the outermost layer is in contact with a contain and is *stationary*. It can also be *turbulent* – there can be lots of internal eddies and swirls of motion, including some that can exist at very small length scales and persist for fair amounts of time. We will idealize all of this – when we discuss static properties of fluids we will assume that all of this sort of internal motion has disappeared.

We can now make a few very simple observations about the forces exerted by the walls of the container on the fluid within. First of all the mass of the fluid in the box above is clearly:

$$\Delta M = \rho \Delta V \tag{8.19}$$

where ΔV is the volume of the box. Since it is at rest and remains at rest, the net external force exerted on it (only) by the the box must be zero (see Week 4). We drew a *symmetric* box to make it easy to see that the magnitudes of the forces exerted by opposing walls are equal $F_{\rm left} = F_{\rm right}$ (for example). Similarly the forces exerted by the top and bottom surfaces, and the front and back surfaces, must cancel.

The average velocity of the molecules in the box must be zero, but the molecules themselves will generally not be at rest at any nonzero temperature. They will be in a state of constant motion where they *bounce elastically* off of the walls of the box, both giving and receiving an impulse (change in momentum) from the walls as they do. The walls of any box large enough to contain many molecules thus exerts a *nearly continuous* nonzero force that confines any fluid not at zero temperature[167].

From this physical picture we can also deduce an important *scaling property* of the force exerted by the walls. We have deliberately omitted giving any actual dimensions to our box in figure 8.2. Suppose (as shown) the cross-sectional area of the left and right walls are ΔA originally. Consider now what we expect if we *double the size* of the box and at the same time add enough additional fluid for the fluid density to remain the same, making the side walls have the area $2\Delta A$. With twice the area (and twice the volume and twice as much fluid), we have twice as many molecular collisions per unit time on the doubled wall areas (with the same average impulse per collision). The average force exerted by the doubled wall areas therefore *also doubles*.

From this simple argument we can conclude that the average force exerted by any wall is *proportional to the area of the wall*. This force is therefore most naturally expressible in terms

[166]This state will also entail *thermodynamic equilibrium* with the box (which must be at a uniform temperature) and hence the fluid in this particular non-accelerating box has a uniform density.

[167]Or at a temperature low enough for the fluid to *freeze* and becomes a *solid*

of *pressure*, for example:

$$F_{\text{left}} = P_{\text{left}}\Delta A = P_{\text{right}}\Delta A = F_{\text{right}} \tag{8.20}$$

which implies that the *pressure* at the left and right confining walls is the same:

$$P_{\text{left}} = P_{\text{right}} = P \tag{8.21}$$

, and that this pressure describes the force exerted by the fluid on the walls and vice versa. Again, the exact same thing is true for the other four sides.

There is nothing special about our particular choice of left and right. If we had originally drawn a *cubic* box (as indeed we did) we can easily see that the pressure P on *all* the faces of the cube must be the same and indeed (as we shall see more explicitly below) the pressure everywhere in the fluid must be the same!

That's quite a lot of mileage to get out of symmetry and the definition of static equilibrium, but there is one more important piece to get. This last bit involves forces exerted by the wall *parallel* to its surface. On average, there cannot be any! To see why, suppose that one surface, say the left one, exerted a force tangent to the surface itself on the fluid in contact with that surface. An important property of fluids is that *one part of a fluid can move independent of another* so the fluid in at least *some* layer with a finite thickness near the wall would therefore experience a *net* force and would *accelerate*. But this violates our assumption of static equilibrium, so a fluid in *static equilibrium* exerts no tangential force on the walls of a confining container and vice versa.

We therefore conclude that the *direction* of the force exerted by a confining surface with an area ΔA on the fluid that is in contact with it is:

$$\vec{\boldsymbol{F}} = P\ \Delta A \hat{\boldsymbol{n}} \tag{8.22}$$

where $\hat{\boldsymbol{n}}$ is an inward-directed unit vector *perpendicular to* (normal to) the surface. This final rule permits us to handle the force exerted on fluids confined to *irregular* amoebic blob shaped containers, or balloons, or bags, or – well, us, by our skins and vascular system.

Note well that this says nothing about the tangential force exerted by fluids in *relative motion* to the walls of the confining container. We already know that a fluid moving across a solid surface will exert a *drag force*, and later this week we'll attempt to at least approximately quantify this.

Next, let's consider what happens when we bring this box of fluid[168] *down to Earth* and consider what happens to the pressure in a box in *near-Earth gravity*.

8.1.7: Pressure and Confinement of Static Fluids in Gravity

The principle change brought about by setting our box of fluid down on the ground in a gravitational field (or equivalently, accelerating the box of fluid uniformly in some direction to develop a *pseudo*-gravitational field in the non-inertial frame of the box) is that an additional external

[168] It's just a box of rain. I don't know who put it there...

force comes into play: The weight of the fluid. A static fluid, confined in some way in a gravitational field, must ***support the weight of its many component parts*** internally, and of course the box itself must support the weight of the entire mass ΔM of the fluid.

As hopefully you can see if you carefully read the previous section. The only force available to provide the necessary internal support or confinement force is the ***variation of pressure within the fluid***. We would like to know how the pressure *varies* as we move up or down in a static fluid so that it supports its own weight.

There are countless reasons that this knowledge is valuable. It is this pressure variation that hurts your ears if you dive deep into the water or collapses submarines if they dive too far. It is this pressure variation that causes your ears to pop as you ride in a car up the side of a mountain or your blood to boil into the vacuum of space if you ride in a rocket all the way out of the atmosphere without a special suit or vehicle that provides a personally pressurized environment. It is this pressure variation that will one day very likely cause you to have varicose veins and edema in your lower extremities from standing on your feet all day – and can help treat/reverse both if you stand in 1.5 meter deep water instead of air. The pressure variation drives water out of the pipes in your home when you open the tap, helps lift a balloon filled with helium, floats a boat but fails to float a rock.

We need to *understand* all of this, whether our eventual goal is to become a physicist, a physician, an engineer, or just a scientifically literate human being. Let's get to it.

Here's the general idea. If we consider a tiny (eventually differentially small) chunk of fluid in force equilibrium, gravity will pull it down and the only thing that can push it up is a *pressure difference* between the top and the bottom of the chunk. By requiring that the force exerted by the pressure difference balance the weight, we will learn how the pressure varies with increasing depth.

For incompressible fluids, this is really all there is to it – it takes only a few lines to derive a lovely formula for the increase in pressure as a function of depth in an incompressible *liquid*.

For gases there is, alas, a small complication. Compressible fluids have densities that *increase* as the pressure increases. This means that boxes of the same size also have *more mass* in them as one descends. More mass means that the pressure difference has to increase, faster, which makes the density/mass greater still, and one discovers (in the end) that the pressure varies *exponentially* with depth. Hence the air pressure drops relatively *quickly* as one goes up from the Earth's surface to very *close* to zero at a height of ten miles, but the atmosphere itself extends for a very long way into space, never quite dropping to "zero" even when one is twenty or a hundred miles high.

As it happens, the calculus for the two kinds of fluids is the *same* up to a given (very important) common point, and then differs, becoming very simple indeed for incompressible fluids and a bit more difficult for compressible ones. Simple solutions suffice to help us build our conceptual understanding; we will therefore treat incompressible fluids first and ***everybody*** is responsible for understanding them. Physics majors, math majors, engineers, and people who love a good bit of calculus now and then should probably continue on and learn how to integrate the simple model provided for compressible fluids.

In figure 8.3 a (portion of) a fluid confined to a box is illustrated. The box could be a completely sealed one with rigid walls on all sides, or it could be something like a cup or

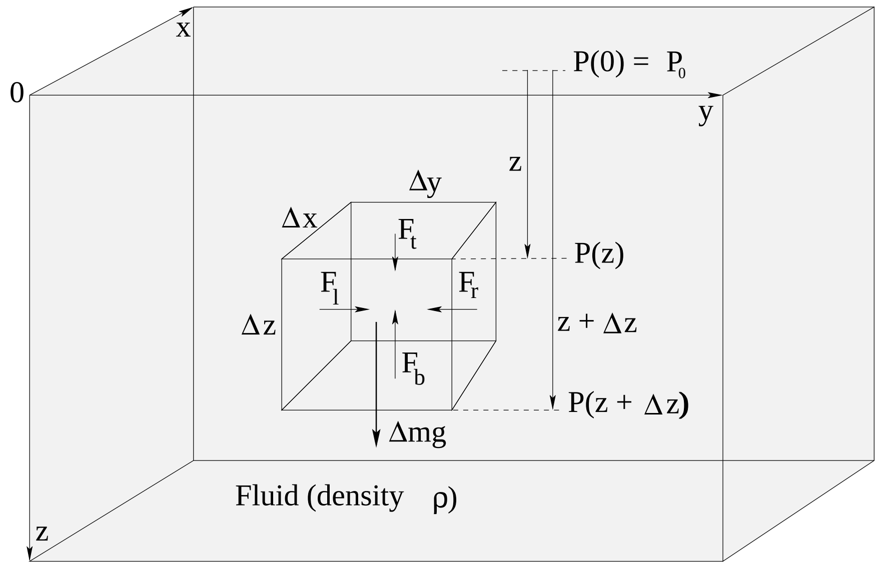


Figure 8.3: A fluid in static equilibrium confined to a sealed rectilinear box in a near-Earth gravitational field $\vec{g}$. Note well the small chunk of fluid with dimensions $\Delta x, \Delta y, \Delta z$ in the middle of the fluid. Also note that the coordinate system selected has z increasing from the top of the box *down*, so that z can be thought of as the *depth* of the fluid.

bucket that is open on the top but where the fluid is still confined there by e.g. atmospheric pressure.

Let us consider a small (eventually infinitesimal) chunk of fluid somewhere in the *middle* of the container. As shown, it has physical dimensions $\Delta x, \Delta y, \Delta z$; its upper surface is a distance z below the origin (where z increases down and hence can represent "depth") and its lower surface is at depth $z + \Delta z$. The areas of the top and bottom surfaces of this small chunk are e.g. $\Delta A_{tb} = \Delta x \Delta y$, the areas of the sides are $\Delta x \Delta z$ and $\Delta y \Delta z$ respectively, and the volume of this small chunk is $\Delta V = \Delta x \Delta y \Delta z$.

This small chunk is itself in static equilibrium – therefore the forces between any pair of its horizontal sides (in the x or y direction) must cancel. As before (for the box in space) $F_l = F_r$ in magnitude (and opposite in their y-direction) and similarly for the force on the front and back faces in the x-direction, which will always be true if the pressure does not vary *horizontally* with variations in x or y. In the z-direction, however, force equilibrium requires that:

$$F_t + \Delta m g - F_b = 0 \tag{8.23}$$

(where recall, down is positive).

The only possible *source* of F_t and F_b are the ***pressure in the fluid itself*** which will ***vary with the depth z***: $F_t = P(z)\Delta A_{tb}$ and $F_b = P(z + \Delta z)\Delta A_{tb}$. Also, the mass of fluid in the (small) box is $\Delta m = \rho \Delta V$ (using our ritual incantation "the mass of the chunks is..."). We can thus write:

$$P(z)\Delta x \Delta y + \rho(\Delta x \Delta y \Delta z)g - P(z + \Delta z)\Delta x \Delta y = 0 \tag{8.24}$$

or (dividing by $\Delta x \Delta y \Delta z$ and rearranging):

$$\frac{\Delta P}{\Delta z} = \frac{P(z + \Delta z) - P(z)}{\Delta z} = \rho g \tag{8.25}$$

Finally, we take the limit $\Delta z \to 0$ and identify the ***definition of the derivative*** to get:

$$\frac{dP}{dz} = \rho g \tag{8.26}$$

Identical arguments but *without* any horizontal external force followed by $\Delta x \to 0$ and $\Delta y \to 0$ lead to:

$$\frac{dP}{dx} = \frac{dP}{dy} = 0 \tag{8.27}$$

as well – P does not vary with x or y as already noted[169].

In order to find $P(z)$ from this differential expression (which applies, recall, to *any* confined fluid in static equilibrium in a gravitational field) we have to *integrate* it. This integral is very simple if the fluid is incompressible because in that case ***ρ is a constant***. The integral isn't *that* difficult if ρ is *not* a constant as implied by the equation we wrote above for the bulk compressibility. We will therefore first do incompressible fluids, then compressible ones.

8.1.8: Variation of Pressure in Incompressible Fluids

In the case of incompressible fluids, ρ is a constant and does not vary with pressure and/or depth. Therefore we can easily multiple $dP/dz = \rho g$ above by dz on both sides and integrate to find P:

$$\begin{aligned} dP &= \rho g \, dz \\ \int dP &= \int \rho g \, dz \\ P(z) &= \rho g z + P_0 \end{aligned} \tag{8.28}$$

where P_0 is the *constant of integration* for both integrals, and practically speaking is the pressure in the fluid at zero depth (wherever that might be in the coordinate system chosen).

Example 8.1.1: Barometers

Mercury barometers were originally invented by Evangelista Torricelli[170] a natural philosopher who acted as Galileo's secretary for the last three months of Galileo's life under house arrest. The invention was inspired by Torricelli's attempt to solve an important engineering problem. The pump makers of the Grand Duke of Tuscany had built powerful pumps intended to raise water twelve or more meters, but discovered that no matter how powerful the pump, water stubbornly refused to rise more than ten meters into a pipe evacuated at the top.

Torricelli demonstrated that a shorter glass tube filled with mercury, when inverted into a dish of mercury, would fall back into a column with a height of roughly 0.76 meters with a vacuum on top, and soon thereafter discovered that the height of the column fluctuated with the pressure of the outside air pressing down on the mercury in the dish, correctly concluding

---

[169] Physics and math majors and other students of multivariate calculus will recognize that I should probably be using partial derivatives here and establishing that $\vec{\nabla} P = \rho \vec{g}$, where in free space we should instead have had $\vec{\nabla} P = 0 \to P$ constant.

[170] Wikipedia: http://www.wikipedia.org/wiki/Evangelista_Torricelli. ,

that water would behave exactly the same way[171]. Torricelli made a number of other important 17th century discoveries, correctly describing the causes of wind and discovering "Torricelli's Law" (an aspect of the Bernoulli Equation we will note below).

In honor of Torricelli, a unit of pressure was named after him. The ***torr*** is the pressure required to push the mercury in Torricelli's barometer up one millimeter. Because mercury barometers were at one time nearly ubiquitous as the most precise way to measure the pressure of the air, a specific height of the mercury column was the original definition of the standard atmosphere. For better or worse, Torricelli's original observation defined one standard atmosphere to be *exactly* "760 millimeters of mercury" (which is a lot to write or say) or as we would now say, "760 torr"[172].

Mercury barometers are now more or less banned, certainly from the workplace, because mercury is a potentially toxic heavy metal. In actual fact, *liquid* mercury is not biologically active and hence is not particularly toxic. Mercury vapor *is* toxic, but the amount of mercury vapor emitted by the exposed surface of a mercury barometer at room temperature is well below the levels considered to be a risk to human health by OSHA unless the barometer is kept in a small, hot, poorly ventilated room with someone who works there over years. This isn't all that common a situation, but with all toxic metals we are probably better safe than sorry[173].

At this point mercury barometers are rapidly disappearing everywhere but from the hands of collectors. Their manufacture is banned in the U.S., Canada, Europe, and many other nations. We had a lovely one (probably more than one, but I recall one) in the Duke Physics Department up until sometime in the 90's[174], but it was – sanely enough – removed and retired during a renovation that also cleaned up most if not all of the asbestos in the building. Ah, my toxic youth...

Still, at one time they were *extremely* common – most ships had one, many households had one, businesses and government agencies had them – knowing the pressure of the air is an important factor in weather prediction. Let's see how they work(ed).

A simple mercury barometer is shown in figure 8.4. It consists of a tube that is completely filled with mercury. Mercury has a specific gravity of 13.534 at a typical room temperature, hence a density of 13534 kg/m^3). The filled tube is then inverted into a small reservoir of mercury (although other designs of the bottom are possible, some with smaller exposed surface area of the mercury). The mercury falls (pulled down by gravity) out of the tube, leaving behind a ***vacuum*** at the top.

We can easily compute the expected height of the mercury column if P_0 is the pressure on

[171] You, too, get to solve Torricelli's problem as one of your homework problems, but armed with a lot better understanding.

[172] Not to be outdone, one standard atmosphere (or atmospheric pressures in weather reporting) in the U.S. is often given as 29.92 barleycorn-derived *inches* of mercury instead of millimeters. Sigh.

[173] The single biggest risk associated with uncontained liquid mercury (in a barometer or otherwise) is that you can easily spill it, and once spilled it is fairly likely to sooner or later make its way into either *mercury vapor* or ***methyl mercury***, both of which are biologically active and highly toxic. Liquid mercury itself you could drink a glass of and it would pretty much pass straight through you with minimal absorption and little to no damage *if* you – um – "collected" it carefully and disposed of it properly on the other side.

[174] I used to work in the small, cramped space with poor circulation where it was located from time to time but never very long at a time and besides, the room was *cold*. But if I seem "mad as a hatter" – mercury nitrate was used in the making of hats and the vapor used to poison the hatters vapor used to poison hat makers – it probably isn't from this...

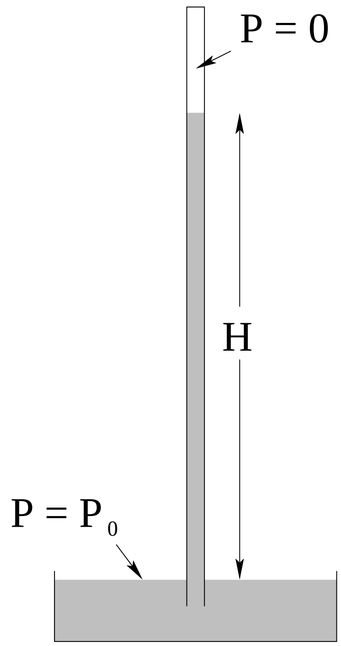


Figure 8.4: A simple fluid barometer consists of a tube with a vacuum at the top filled with fluid supported by the air pressure outside.

the exposed surface of the mercury in the reservoir. In that case

$$P = P_0 + \rho g z \tag{8.29}$$

as usual for an incompressible fluid. Applying this formula to both the top and the bottom,

$$P(0) = P_0 \tag{8.30}$$

and

$$P(H) = P_0 - \rho g H = 0 \tag{8.31}$$

(recall that the upper surface is *above* the lower one, $z = -H$). From this last equation:

$$P_0 = \rho g H \tag{8.32}$$

and one can easily convert the measured height H of mercury above the top surface of mercury in the reservoir into P_0, the air pressure on the top of the reservoir.

At one standard atmosphere, we can easily determine what a mercury barometer at room temperature will read (the height H of its column of mercury above the level of mercury in the reservoir):

$$P_0 = 13534\frac{\text{kg}}{\text{m}^3} \times 9.80665\frac{\text{m}}{\text{sec}^2} \times H = 101325\text{Pa} \tag{8.33}$$

Note well, we have used the precise SI value of g in this expression, and the density of mercury at "room temperature" around 20°C or 293°K. Dividing, we find the expected height of mercury in a barometer at room temperature and one standard atmosphere is $H = 0.763$ meters or 763 torr

Note that this is not exactly the 760 torr we expect to read for a standard atmosphere. This is because for high precision work one cannot just use *any old* temperature (because mercury has a significant thermal expansion coefficient and was then and continues to be used today in mercury thermometers as a consequence). The unit definition is based on using the density of mercury at 0°C or 273.16°K, which has a specific gravity (according to NIST, the National Institute of Standards in the US) of 13.595. Then the precise connection between SI units and torr follows from:

$$P_0 = 13595 \frac{\mathrm{kg}}{\mathrm{m}^3} \times 9.80665 \frac{\mathrm{m}}{\mathrm{sec}^2} \times H = 101325\mathrm{Pa} \tag{8.34}$$

Dividing we find the value of H expected at one standard atmosphere:

$$H_{\mathrm{atm}} = 0.76000 = 760.00 \text{ millimeters} \tag{8.35}$$

Note well the precision, indicative of the fact that the SI units for a standard atmosphere *follow* from their definition in torr, not the other way around.

Curiously, this value is invariably given in both textbooks and even the wikipedia article on atmospheric pressure as the *average* atmospheric pressure at sea level, which it almost certainly is not – a spatiotemporal averaging of sea level pressure would have been utterly impossible during Torricelli's time (and would be difficult today!) and if it was done, could not possibly have worked out to be *exactly* 760.00 millimeters of mercury at 273.16°K.

Example 8.1.2: Variation of Oceanic Pressure with Depth

The pressure on the surface of the ocean is, approximately, by definition, one atmosphere. Water is a highly incompressible fluid with $\rho_w = 1000$ kilograms per cubic meter[175]. $g \approx 10$ meters/second^2. Thus:

$$P(z) = P_0 + \rho_w g z = \left(10^5 + 10^4 z\right) \ \mathrm{Pa} \tag{8.36}$$

or

$$P(z) = (1.0 + 0.1z) \ \mathrm{bar} = (1000 + 100z) \ \mathrm{mbar} \tag{8.37}$$

Every ten meters of depth (either way) increases water pressure by (approximately) one atmosphere!

Wow, *that* was easy. This is a *very important rule of thumb* and is actually fairly easy to remember! How about compressible fluids?

8.1.9: Variation of Pressure in Compressible Fluids

Compressible fluids, as noted, have a density which *varies* with pressure. Recall our equation for the compressibility:

$$\Delta P = -B \frac{\Delta V}{V} \tag{8.38}$$

If one increases the pressure, one therefore *decreases* occupied volume of any given chunk of mass, and hence increases the density. However, to predict precisely how the density will depend on pressure requires more than just this – it requires a *model* relating pressure, volume and mass.

[175] Good number to remember. In fact, *great* number to remember.

Just such a model for a compressible gas is provided (for example) by the ***Ideal Gas Law***[176] :

$$PV = Nk_bT = nRT \tag{8.39}$$

where N is the number of molecules in the volume V, k_b is Boltzmann's constant[177] n is the number of moles of gas in the volume V, R is the ideal gas constant[178] and T is the temperature in degrees Kelvin (or Absolute)[179] . If we assume constant temperature, and convert N to the mass of the gas by multiplying by the molar mass and dividing by Avogadro's Number[180] 6×10^{23}.

(***Aside:*** If you've never taken chemistry a lot of this is going to sound like Martian to you. Sorry about that. As always, consider visting the e.g. Wikipedia pages linked above to learn enough about these topics to get by for the moment, or just keep reading as the *details* of all of this won't turn out to be very important...)

When we do this, we get the following formula for the density of an ideal gas:

$$\rho = \frac{M}{RT}P \tag{8.40}$$

where M is the molar mass[181] , the number of kilograms of the gas per mole. Note well that this result is *idealized* – that's why they call it the *Ideal* Gas Law! – and that no real gases are "ideal" for all pressures and temperatures because sooner or later they all become *liquids* or *solids* due to molecular interactions. However, the gases that make up "air" are all reasonably ideal at temperatures in the ballpark of room temperature, and in any event it is worth seeing how the pressure of an ideal gas varies with z to get an *idea* of how air pressure will vary with height. Nature will probably be somewhat different than this prediction, but we ought to be able to make a *qualitatively* accurate model that is also moderately *quantitatively* predictive as well.

As mentioned above, the formula for the derivative of pressure with z is unchanged for compressible or incompressible fluids. If we take $dP/dz = \rho g$ and multiply both sides by dz as before and integrate, now we get (assuming a fixed temperature T):

$$\begin{aligned}
dP &= \rho g\, dz = \frac{Mg}{RT}P\, dz \\
\frac{dP}{P} &= \frac{Mg}{RT}dz \\
\int \frac{dP}{P} &= \frac{Mg}{RT}\int dz \\
\ln(P) &= \frac{Mg}{RT}z + C
\end{aligned}$$

We now do the usual[182] – exponentiate both sides, turn the exponential of the sum into the product of exponentials, turn the exponential of a constant of integration into a constant of integration, and match the units:

$$P(z) = P_0 e^{\left(\frac{Mg}{RT}\right)z} \tag{8.41}$$

[176]Wikipedia: http://www.wikipedia.org/wiki/Ideal Gas Law.
[177]Wikipedia: http://www.wikipedia.org/wiki/Boltzmann's Constant. ,
[178]Wikipedia: http://www.wikipedia.org/wiki/Gas Constant. ,
[179]Wikipedia: http://www.wikipedia.org/wiki/Temperature.
[180]Wikipedia: http://www.wikipedia.org/wiki/Avogadro's Number.
[181]Wikipedia: http://www.wikipedia.org/wiki/Molar Mass.
[182]Which should be familiar to you both from solving the linear drag problem in Week 2 *and* from the online Math Review.

where P_0 is the pressure at zero *depth*, because (recall!) z is measured positive *down* in our expression for dP/dz.

Example 8.1.3: Variation of Atmospheric Pressure with Height

Using z to describe depth is moderately inconvenient, so let us define the ***height h above sea level*** to be $-z$. In that case P_0 is (how about that!) *1 Atmosphere*. The molar mass of dry air is $M = 0.029$ kilograms per mole. $R = 8.31$ Joules/(mole-K°). Hence a bit of multiplication at $T = 300°$:

$$\frac{Mg}{RT} = \frac{0.029 \times 10}{8.31 \times 300} = 1.12 \times 10^{-4} \text{ meters}^{-1} \quad (8.42)$$

or:

$$P(h) = 10^5 \exp(-0.00012\ h) \text{ Pa} = 1000 \exp(-0.00012\ h) \text{ mbar} \quad (8.43)$$

Note well that the temperature of air is *not* constant as one ascends – it drops by a fairly significant amount, even on the absolute scale (and higher still, it *rises* by an even *greater* amount before dropping again as one moves through the layers of the atmosphere. Since the pressure is found from an integral, this in turn means that the exponential behavior itself is rather inexact, but still it isn't a *terrible* predictor of the variation of pressure with height. This equation predicts that air pressure should drop to $1/e$ of its sea-level value of 1000 mbar at a height of around 8000 meters, the height of the so-called ***death zone***[183] . We can compare the actual (average) pressure at 8000 meters, 356 mbar, to $1000 \times e^{-1} = 368$ mbar. We get remarkably good agreement!

This agreement rapidly breaks down, however, and meteorologists actually use a patchwork of formulae (both algebraic and exponential) to give better agreement to the actual variation of air pressure with height as one moves up and down through the various named layers of the atmosphere with the pressure, temperature and even molecular composition of "air" varying all the way. This simple model explains a *lot* of the variation, but its assumptions are not really correct.

8.2: Pascal's Principle and Hydraulics

We note that (from the above) the *general form* of P of a fluid confined to a sealed container has the most general form:

$$P(z) = P_0 + \int_0^z \rho g dz \quad (8.44)$$

where P_0 is the constant of integration or value of the pressure at the reference ***depth*** $z = 0$. This has an important consequence that forms the basis of *hydraulics*.

[183] Wikipedia: http://www.wikipedia.org/wiki/Effects of high altitude on humans. This is the height where air pressure drops to where humans are at extreme risk of dying if they climb without supplemental oxygen support – beyond this height hypoxia reduces one's ability to make important and life-critical decisions during the very last, most stressful, part of the climb. Mount Everest (for example) can only be climbed with oxygen masks and some of the greatest disasters that have occurred climbing it and other peaks are associated with a lack of or failure of supplemental oxygen.

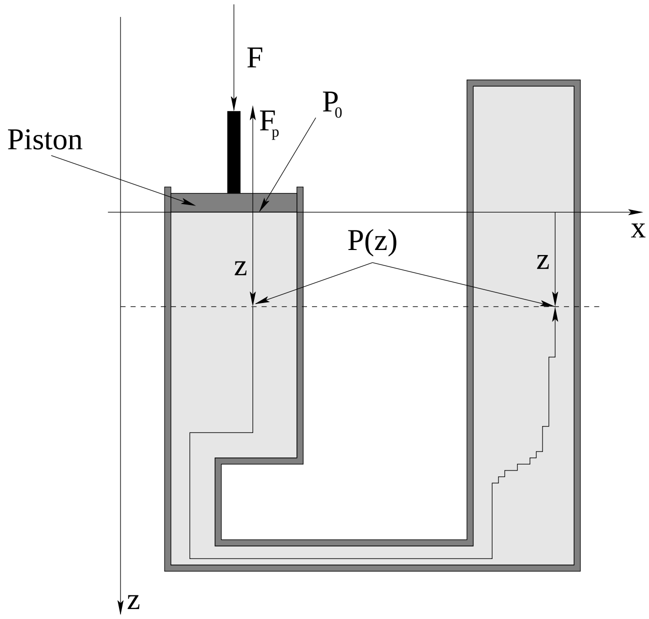


Figure 8.5: A single piston seated tightly in a frictionless cylinder of cross-sectional area A is used to compress water in a sealed container. Water is incompressible and does not significantly change its volume at $P = 1$ bar (and a constant room temperature) for pressure changes on the order of 0.1-100 bar.

Suppose, then, that we have an ***incompressible fluid*** e.g. water confined within a sealed container by e.g. a *piston* that can be pushed or pulled on to *increase or decrease* the *confinement pressure* on the surface of the piston. Such an arrangement is portrayed in figure 8.5.

We can push down (or pull back) on the piston with any total downward force F that we like that leaves the system in equilibrium. Since the piston itself is in static equilibrium, the force we push with must be opposed by the pressure in the fluid, which exerts an equal and opposite upwards force:

$$F = F_p = P_0 A \tag{8.45}$$

where A is the cross sectional area of the piston and where we've put the cylinder face at $z = 0$, which we are obviously free to do. For all practical purposes this means that we can make P_0 "anything we like" within the range of pressures that are unlikely to make water at room temperature change it's state or volume do other bad things, say $P = (0.1, 100)$ bar.

The pressure at a depth z in the container is then (from our previous work):

$$P(z) = P_0 + \rho g z \tag{8.46}$$

where $\rho = \rho_w$ if the cylinder is indeed filled with water, but the cylinder could equally well be filled with hydraulic fluid (basically oil, which assists in lubricating the piston and ensuring that it remains "frictionless' while assisting the seal), alcohol, mercury, or any other incompressible liquid.

We recall that the pressure changes *only* when we change our depth. Moving laterally does not change the pressure, because e.g. $dP/dx = dP/dy = 0$. We can always find a path consisting of vertical and lateral displacements from $z = 0$ to any other point in the container

– two such points at the same depth z are shown in **??**, along with a (deliberately ziggy-zaggy[184]) vertical/horizontal path connecting them. Clearly these two points ***must have the same pressure*** $\boldsymbol{P(z)}$***!***

Now consider the following. Suppose we start with pressure P_0 (so that the pressure at these two points is $P(z)$, but then change F to make the pressure P_0' and the pressure at the two points $P'(z)$. Then:

$$\begin{aligned} P(z) &= P_0 + \rho g z \\ P'(z) &= P_0' + \rho g z \\ \Delta P(z) &= P'(z) - P(z) = P_0' - P_0 = \Delta P_0 \end{aligned} \tag{8.47}$$

That is, the pressure change at depth z does not depend on z *at any point in the fluid!* It depends *only on the change in the pressure exerted by the piston!*

This result is known as ***Pascal's Principle*** and it holds (more or less) for *any* compressible fluid, not just incompressible ones, but in the case of compressible fluids the piston will move up or down or in or out and the density of the fluid will change and hence the treatment of the integral will be too complicated to cope with. Pascal's Principle is more commonly given in *English words* as:

> ***Any change in the pressure exerted at a given point on a confined fluid is transmitted, undiminished, throughout the fluid.***

Pascal's principle is the basis of ***hydraulics***. Hydraulics are a kind of fluid-based simple machine that can be used to greatly amplify an applied force. To understand it, consider the following figure:

Example 8.2.1: A Hydraulic Lift

Figure 8.6 illustrates the way we can multiply forces using Pascal's Principle. Two pistons seal off a pair of cylinders connected by a closed tube that contains an incompressible fluid. The two pistons are deliberately given the same height (which might as well be $z = 0$, then, in the figure, although we could easily deal with the variation of pressure associated with them being at different heights since we know $P(z) = P_0 + \rho g z$. The two pistons have cross sectional areas A_1 and A_2 respectively, and support a small mass m on the left and large mass M on the right in static equilibrium.

For them to be in equilibrium, clearly:

$$F_1 - mg = 0 \tag{8.48}$$

$$F_2 - Mg = 0 \tag{8.49}$$

We also/therefore have:

$$F_1 = P_0 A_1 = mg \tag{8.50}$$

$$F_2 = P_0 A_2 = Mg \tag{8.51}$$

[184]Because we can make the zigs and zags *differentially small*, at which point this piecewise horizontal-vertical line becomes an *arbitrary curve* that remain in the fluid. Multivariate calculus can be used to formulate all of these results more prettily, but the *reasoning* behind them is completely contained in the picture and this text explanation.

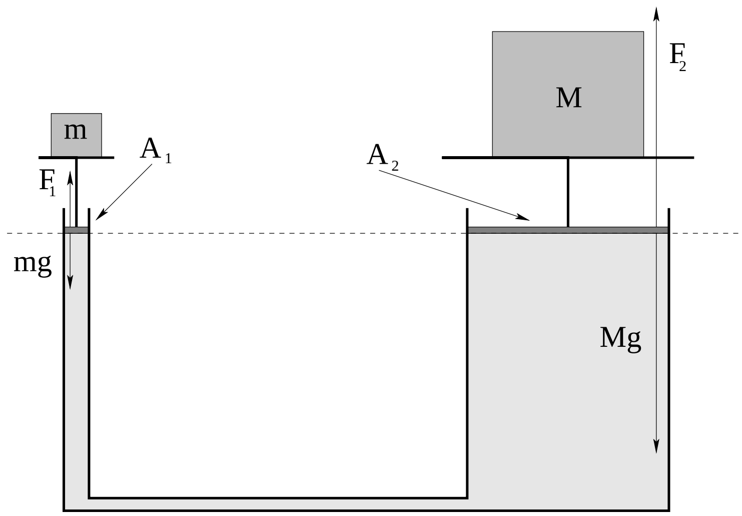


Figure 8.6: A simple schematic for a hydraulic lift of the sort used in auto shops to lift your car.

Thus

$$\frac{F_1}{A_1} = P_0 = \frac{F_2}{A_2} \tag{8.52}$$

or (substituting and cancelling g):

$$M = \frac{A_2}{A_1} m \tag{8.53}$$

A small mass on a small-area piston can easily balance a *much larger mass on an equally larger area piston!*

Just like a lever, we can balance or lift a large weight with a small one. Also just as was the case with a lever, *there ain't no such thing as a free lunch!* If we try to *lift* (say) a car with a hydraulic lift, we have to move the same volume $\Delta V = A\Delta z$ from under the small piston (as it descends) to under the large one (as it ascends). If the small one goes down a distance z_1 and the large one goes up a distance z_2, then:

$$\frac{z_1}{z_2} = \frac{A_2}{A_1} \tag{8.54}$$

The *work* done by the two cylinders thus *precisely balances*:

$$W_2 = F_2 z_2 = F_1 \frac{A_2}{A_1} z_2 = F_1 \frac{A_2}{A_1} z_1 \frac{A_1}{A_2} = F_1 z_1 = W_1 \tag{8.55}$$

The hydraulic arrangement thus transforms pushing a small force through a large distance into a large force moved through a small distance so that the work done ***on*** piston 1 matches the work done ***by*** piston 2. No energy is created or destroyed (although in the real world a bit will be lost to heat as things move around) and all is well, quite literally, with the Universe.

This example is pretty simple, but it should suffice to guide you through doing a work-energy conservation problem where (for example) the mass m goes down a distance d (losing gravitational energy) and the mass M goes up a distance D (gaining gravitational energy *while the fluid itself also is net moved up above its former level!* Don't forget that last, tricky bit if you ever have a problem like that!

8.3: Fluid Displacement and Buoyancy

First, a story. Archimedes[185] was, quite possibly, the smartest person who has ever lived (so far). His day job was being the "court magician" in the island kingdom of Syracuse in the third century BCE, some 2200 years ago; in his free time he did things like invent primitive integration, accurately compute pi, invent amazing machines of war and peace, determine the key principles of both statics and fluid statics (including the one we are about to study and the principles of the lever – "Give me put a place to stand and I can move the world!" is a famous Archimedes quote, implying that a sufficiently long lever would allow the small forces humans can exert to move even something as large as the Earth, although yeah, there are a few problems with that that go beyond just a place to stand[186]).

The king (Hiero II) of Syracuse had a problem. He had given a goldsmith a mass of pure gold to make him a ***votive crown***, but when the crown came back he had the niggling suspicion that the goldsmith had substituted cheap silver for some of the gold and kept the gold. It was keeping him awake at nights, because if somebody can steal from the king and get away with it (and word gets out) it can only encourage a loss of respect and rebellion.

So he called in his court magician (Archimedes) and gave him the task of determining whether or not the crown had been made by adulterated gold – or else. And oh, yeah – you can't melt down the crown and cast it back into a regular shape whose dimensions can be directly compared to the same shape of gold, permitting a direct comparison of their *densities* (the density of pure gold is not equal to the density of gold with an admixture of silver). And don't forget the "or else".

Archimedes puzzled over this for some days, and decided to take a bath and cool off his overheated brain. In those days, baths were large public affairs – you *went* to the baths as opposed to having one in your home – where a filled tub was provided, sometimes with attendants happy to help you wash. As the possibly apocryphal story has it, Archimedes lowered himself into the overfull tub and as he did so, water sloshed out as he *displaced its volume* with his own volume. In an intuitive, instantaneous flash of insight – a "light bulb moment" – he realized that *displacement of a liquid by an irregular shaped solid* can be used to measure its volume, and that such a measurement of displaced volume would allow the king's problem to be solved.

Archimedes then leaped out of the tub and ran naked through the streets of Syracuse (which we can only imagine provided its inhabitants with as much amusement then as it would provide now) yelling "Eureka!", which in Greek means "I have found it!" The test (two possible versions of which are supplied below, one more probable than the other but less instructive for our own purposes) was performed, and revealed that the goldsmith was indeed dishonest and had stolen some of the king's gold. Bad move, goldsmith! We will draw a tasteful veil over the probable painful and messy fate of the goldsmith.

---

[185] Wikipedia: http://www.wikipedia.org/wiki/Archimedes. A very, very interesting person. I strongly recommend that my students read this short article on this person who came *within a hair* of inventing physics and calculus and starting the Enlightenment some 1900 years before Newton. Scary supergenius polymath guy. Would have won multiple Nobel prizes, a Macarthur "Genius" grant, and so on if alive today. Arguably the smartest person who has ever lived – so far.

[186] The "sound bite" is hardly a modern invention, after all. Humans have always loved a good, pithy statement of insight, even if it isn't actually even approximately true...

Archimedes transformed his serendipitous discovery of static fluid displacement into an elaborate physical principle that explained *buoyancy*, the tendency of fluids to support all or part of the weight of objects immersed in them.

The fate of Archimedes himself is worth a moment more of our time. In roughly 212 BCE, the Romans invaded Syracuse in the Second Punic War after a two year siege. As legend has it, as the city fell and armed soldiers raced through the streets "subduing" the population as only soldiers can, Archimedes was in his court chambers working on a problem in the geometry of circles, which he had drawn out in the sand boxes that then served as a "chalkboard". A Roman soldier demanded that he leave his work and come meet with the conquering general, Marcus Claudius Marcellus. Archimedes declined, replying with his last words "Do not disturb my circles" and the soldier killed him. Bad move, soldier – Archimedes himself was a major part of the loot of the city and Marcellus had ordered that he was not to be harmed. The fate of the soldier that killed him is unknown, but it wasn't really a very good idea to anger a conquering general by destroying an object or person of enormous value, and I doubt that it was very good.

Anyway, let's see the modern version of Archimedes' discovery and see as well how Archimedes very probably used it to test the crown.

8.3.1: Archimedes' Principle

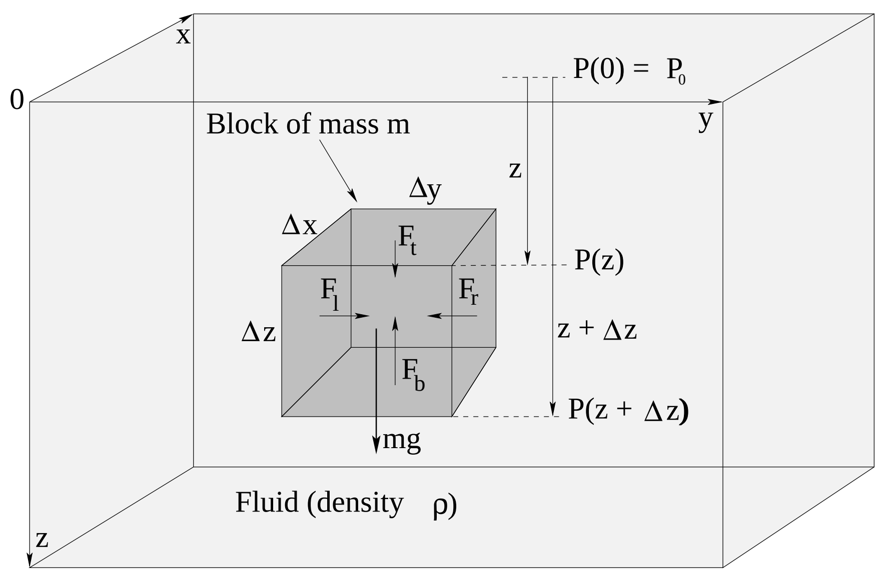


Figure 8.7: A solid chunk of "stuff" of mass m and the dimensions shown is immersed in a fluid of density ρ at a depth z. The vertical pressure difference in the fluid (that arises as the fluid itself becomes static static) exerts a vertical force on the cube.

If you are astute, you will note that figure 8.7 is ***exactly like figure 8.3*** above, except that the internal chunk of *fluid* has been replace by *some other material*. The point is that this replacemend does not matter – ***the net force exerted on the cube by the fluid is the same!***

Hopefully, that is obvious. The net upward force exerted by the fluid is called the ***buoyant***

force F_b and is equal to:

$$\begin{aligned}
F_b &= P(z+\Delta z)\Delta x\Delta y - P(z)\Delta x\Delta y \\
&= \left((P_0+\rho g(z+\Delta z)) - (P_0+\rho g z)\right)\Delta x\Delta y \\
&= \rho g\Delta z\Delta x\Delta y \\
&= \rho g\Delta V \qquad (8.56)
\end{aligned}$$

where ΔV is the volume of the small block.

The buoyant force is thus the ***weight of the fluid displaced*** by this single tiny block. This is all we need to show that the same thing is true for an *arbitrary* immersed shape of object.

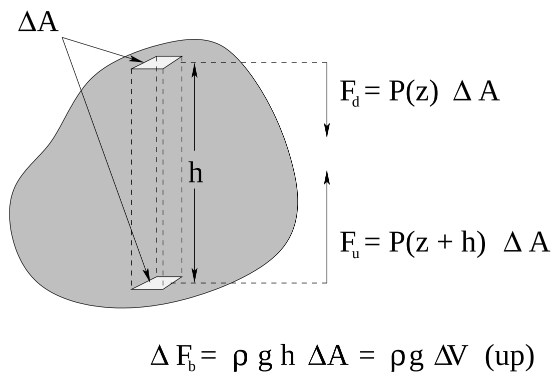


Figure 8.8: An arbitrary chunk of stuff is immersed in a fluid and we consider a vertical cross section with horizontal ends of area ΔA and height h through the chunk.

In figure 8.8, an arbitrary blob-shape is immersed in a fluid (not shown) of density ρ. Imagine that we've taken a french-fry cutter and cut the whole blob into nice rectangular segments, one of which (of length h and cross-sectional area ΔA) is shown. We can trim or average the end caps so that they are all perfectly horizontal by making all of the rectangles arbitrarily small (in fact, differentially small in a moment). In that case the *vertical* force exerted by the fluid on just the two lightly shaded surfaces shown would be:

$$F_d = P(z)\Delta A \qquad (8.57)$$

$$F_u = P(z+h)\Delta A \qquad (8.58)$$

where we assume the upper surface is at depth z (this won't matter, as we'll see in a moment). Since $P(z+h) = P(z) + \rho g h$, we can find the *net upward buoyant force* exerted on this little cross-section by subtracting the first from the second:

$$\Delta F_b = F_u - F_d = \rho g\, h\Delta A = \rho g\, \Delta V \qquad (8.59)$$

where the *volume* of this piece of the entire blob is $\Delta V = h\,\Delta A$.

We can now let $\Delta A \to dA$, so that $\Delta V -> dV$, and write

$$F_b = \int dF_b = \int_{V \text{ of blob}} \rho g\, dV = \rho g V = m_f g \qquad (8.60)$$

where $m_f = \rho V$ is the mass of the fluid displaced, so that $m_f g$ is its weight.

That is:

> ***The total buoyant force on the immersed object is the weight of the fluid displaced by the object.***

This is really an adequate *proof* of this statement, although if we were really going to be picky we'd use the fact that P doesn't vary in x or y to show that the net force in the x or y direction is zero independent of the shape of the blob, using our differential french-fry cutter mentally in the x direction and then noting that the blob is *arbitrary* in shape and we could have just as easily labelled or oriented the blob with this direction called y so it must be true in *any* direction perpendicular to $\vec{g}$.

This statement – in the English or algebraic statement as you prefer – is known as ***Archimedes' Principle***, although Archimedes could hardly have formulated it quite the way we did algebraically above as he died before he could *quite* finish inventing the calculus and physics.

This principle is enormously important and ubiquitous. Buoyancy is why boats float, but rocks don't. It is why childrens' helium-filled balloons do odd things in accelerating cars. It exerts a subtle force on everything submerged in the air, in water, in beer, in liquid mercury, as long as the fluid itself is either in a gravitational field (and hence has a pressure gradient) or is in an accelerating container with its own "pseudogravity" (and hence has a pressure gradient).

Let's see how Archimedes could have used this principle to test the crown two ways. The first way is very simple and *conceptually* instructive; the second way is more practical to us as it illustrates the way we generally do algebra associated with buoyancy problems.

Example 8.3.1: Testing the Crown I

The tools Archimedes probably had available to him were balance-type scales, as these tools for comparatively measuring the weight were well-known in antiquity. He certainly had vessels he could fill with water. He had thread or string, he had the crown itself, and he had access to pure gold from the king's treasury (at least for the duration of the test.

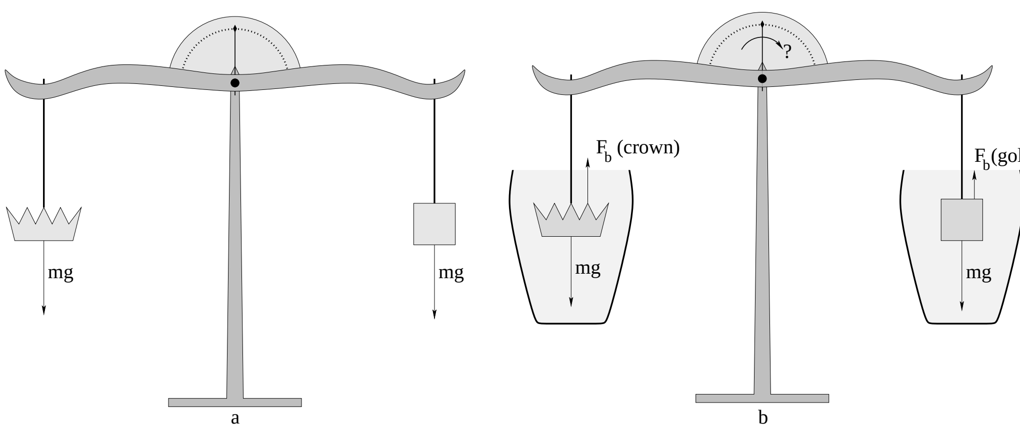


Figure 8.9: In a), the crown is balanced against an equal weight/mass of pure gold *in air*. In b) the crown *and* the gold are symmetrically submerged in containers of still water.

Archimedes very likely used his balance to first select and trim a piece of gold so it had *exactly the same weight as the crown* as illustrated in figure 8.9a. Then all he had to do was submerge the crown and the gold symmetrically in two urns filled with water, taking care that they are both fully underwater.

Pure gold is *more dense* than gold adulterated with silver (the most likely metal the goldsmith would have used; although a few others such as copper might have also been available and/or used they are also less dense than gold). This means that any given mass/weight (in air, with its negligible buoyant force) of adulterated gold would have a ***greater volume*** than an equal mass/weight (in air) of pure gold.

If the crown were made of pure gold, then, the buoyant forces acting on the gold and the crown would be equal. The weights of the gold and crown are equal. Therefore the submerged crown and submerged gold would be supported in static equilibrium by the *same force* on the ropes, and the balance would indicate "equal" (the indicator needle straight up). The goldsmith lives, the king is happy, Archimedes lives, everybody is happy.

If the crown is made of less-dense gold *alloy*, then its volume will be greater than that of pure gold. The buoyant force acting on it when submerged will therefore *also* be greater, so the tension in the string supporting it needed to keep it in static equilibrium will be smaller.

But this smaller tension then would fail to balance the *torque* exerted on the balance arms by the string attached to the gold, and the whole balance would rotate to the right, with the more dense gold sinking relative to the less dense crown. The balance needle would *not* read "equal". In the story, it didn't read equal. So sad – for the goldsmith.

Example 8.3.2: Testing the Crown II

Of course nowadays we don't do things with balance-type scales so often. More often than not we would use a *spring balance* to weigh something from a string. The good thing about a spring balance is that you can *directly read off the weight* instead of having to delicately balance some force or weight with masses in a counterbalance pan. Using such a balance (or any other accurate scale) we can measure and record the *density of pure gold* once and for all.

Let us imagine that we have done so, and discovered that:

$$\rho_{\rm Au} = 19300 \text{ kilograms/meter}^3 \tag{8.61}$$

For grins, please note that $\rho_{\rm Ag} = 10490$ kilograms/meter^3. This is a bit over half the density of gold, so that adulterating the gold of the crown with 10% silver would have decreased its density by around 5%. If the mass of the crown was (say) a kilogram, the goldsmith would have stolen 100 grams – almost four ounces – of pure gold at the cost of 100 grams of silver. Even if he stole twice that, the 9% increase in volume would have been nearly impossible to directly observe in a worked piece. At that point the *color* of the gold would have been off, though. This could be remedied by adding *copper* ($\rho_{\rm Cu} = 8940$ kilograms/meter^3) along with the silver. Gold-Silver-Copper all three alloy together, with silver whitening and yellowing the natural color of pure gold and copper reddening it, but with the two *balanced* one can create an alloy that is perhaps 10% *each* copper and silver that has almost exactly the same color as pure gold. This would *harden* and *strengthen* the gold of the crown, but you'd have to damage the crown to discover this.

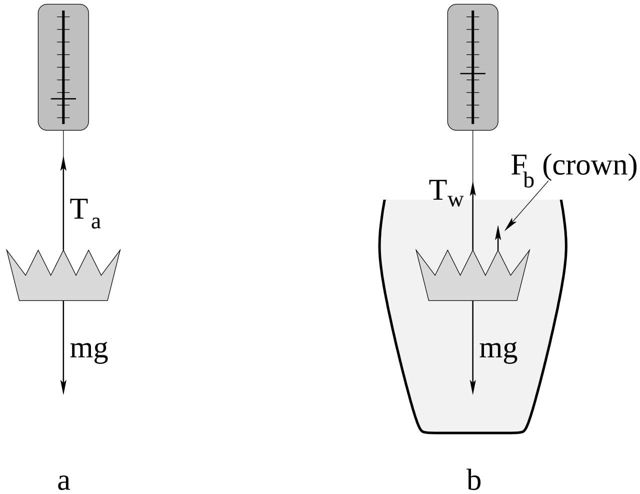


Figure 8.10: We now measure the effective weight of the *one* crown both in air (very close to its true weight) and in water, where the measured weight is *reduced* by the buoyant force.

Instead we hang the crown (of mass m) as before, but this time from a spring balance, both in air and in the water, recording both weights *as measured by the balance* (which measures, recall, the tension in the supporting string). This is illustrated in figure 8.10, where we note that the measured weights in a) and b) are T_a, the weight in air, and T_w, the measured weight while immersed in water.

Let's work this out. a) is simple. In static equilibrium:

$$\begin{aligned} T_a - mg &= 0 \\ T_a &= mg \\ T_a &= \rho_{\rm crown} V g \end{aligned} \tag{8.62}$$

so the scale in a) just measures the almost-true weight of the crown (off by the buoyant force exerted by the *air* which, because the density of air is very small at $\rho_{\rm air} \approx 1.2$ kilograms/meter^3, which represents around a 0.1% error in the measured weight of objects roughly the density of water, and an even smaller error for denser stuff like gold).

In b):

$$\begin{aligned} T_w + F_b - mg &= 0 \\ T_w &= mg - F_b \\ &= \rho_{\rm crown} V g - \rho_{\rm water} V g \\ &= (\rho_{\rm crown} - \rho_{\rm water}) V g \end{aligned} \tag{8.63}$$

We know (we *measured*) the values of T_a and T_w, but we don't know V or $\rho_{\rm crown}$. We have two equations and two unknowns, and we would like most of all to solve for $\rho_{\rm crown}$. To do so, we *divide these two equations by one another* to eliminate the V:

$$\frac{T_w}{T_a} = \frac{\rho_{\rm crown} - \rho_{\rm water}}{\rho_{\rm crown}} \tag{8.64}$$

Whoa! g went away too! This means that from here on we don't even care what g is – we could make these weight measurements on the moon or on mars and we'd still get the *relative* density of the crown (compared to the density of water) right!

A bit of algebra-fu:

$$\rho_{\rm water} = \rho_{\rm crown}(1 - \frac{T_w}{T_a}) = \rho_{\rm crown}\frac{T_a - T_w}{T_a} \tag{8.65}$$

or finally:

$$\rho_{\rm crown} = \rho_{\rm water}\frac{T_a}{T_a - T_w} \tag{8.66}$$

We are now prepared to be precise. Suppose that the color of the crown is very good. We perform the measurements above (using a scale accurate to *better* than a hundredth of a Newton or we might end up condemning our goldsmith due to a *measurement error!*) and find that $T_a = 10.00$ Newtons, $T_w = 9.45$ Newtons. Then

$$\rho_{\rm crown} = 1000\frac{10.00}{10.00 - 9.45} = 18182 \text{ kilograms/meter}^3 \tag{8.67}$$

We subtract, $19300 - 18182 = 1118$; divide, $1118/19300 \times 100 = 6$%. Our crown's material is around *six percent less dense than gold* which means that our clever goldsmith has adulterated the gold by removing some 12% of the gold (give or take a percent) and replaced it with some mixture of silver and copper. Baaaaad goldsmith, bad.

If the goldsmith were smart, of course, he could have beaten Archimedes (and us). What he needed to do is adulterate the gold with a mixture of metals that have *exactly the same density as gold!* Not so easy to do, but tungsten's density, $\rho_{\rm W} = 19300$ (to three digits) almost exactly matches that of gold. Alas, it has the highest melting point of all metals at 3684°K, is enormously hard, and might or might not alloy with gold or change the color of the gold if alloyed. It is also pretty expensive in its own right. Platinum, Plutonium, Iridium, and Osmium are all even denser then gold, but three of these are very expensive (even more expensive than gold!) and one is very explosive, a transuranic compound used to make nuclear bombs, enormously expensive *and* illegal to manufacture or own (and rather toxic as well). Not so easy, matching the density via adulteration and making a profit out of it...

Enough of all of this *fluid statics*. Time to return to some *dynamics*.

8.4: Fluid Flow

In figure 8.11 we see fluid flowing from left to right in a circular pipe. The pipe is assumed to be "frictionless" for the time being – to exert no drag force on the fluid flowing within – and hence all of the fluid is moving *uniformly* (at the same speed v with no relative *internal* motion) in a state of ***dynamic equilibrium***.

We are interested in understanding the ***flow*** or ***current*** of water carried by the pipe, which we will define to be the ***volume per unit time*** that passes any given point in the pipe. Note well that we could instead talk about the *mass per unit time* that passes a point, but this is just the volume per unit time times the density and hence for fluids with a more or less *uniform* density the two are the same within a constant.

For this reason we will restrict our discussion in the following to incompressible fluids, with constant ρ. This means that the concepts we develop will work gangbusters well for understanding water flowing in pipes, beer flowing from kegs, blood flowing in veins, and even rivers

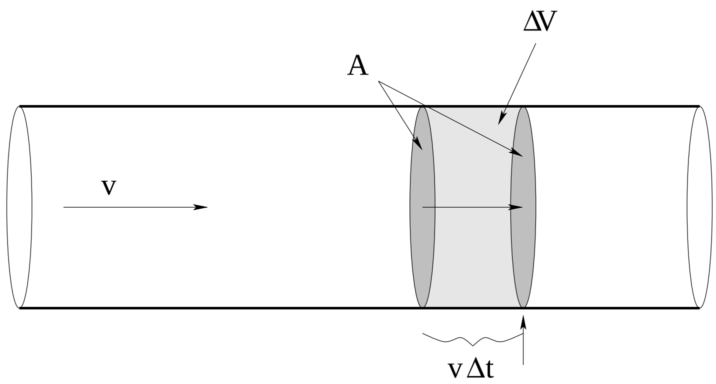


Figure 8.11: Fluid in uniform flow is transported down a pipe with a constant cross-section at a constant speed v. From this we can easily compute the *flow*, the volume per unit time that passes (through a surface that cuts the pipe at) a point on the pipe.

flowing slowly in not-too-rocky river beds but not so well to describe the dynamical evolution of weather patterns or the movement of oceanic currents. The *ideas* will still be extensible, but future climatologists or oceanographers will have to work a bit harder to understand the correct theory when dealing the compressibility.

We expect a "big pipe" (one with a large cross-sectional area) to carry more fluid per unit time, all other things being equal, than a "small pipe". To understand the relationship between area, speed and flow we turn our attention to figure **??**. In a time Δt, all of the water within a distance $v\Delta t$ to the left of the second shaded surface (which is strictly imaginary – there is nothing actually in that pipe at that point but fluid) will pass *through* this surface and hence past the point indicated by the arrow underneath. The volume of this fluid is just the area of the surface times the height of the cylinder of water:

$$\Delta V = Av\Delta t \tag{8.68}$$

If we divide out the Δt, we get:

$$I = \frac{\Delta V}{\Delta t} = Av \tag{8.69}$$

This, then is the ***flow***, or ***volumetric current*** of fluid in the pipe.

This is an *extremely important relation*, but the picture and derivation itself is arguably even *more* important, as this is the first time – but ***not the last time*** – you have seen it, and it will be a crucial part of understanding things like ***flux*** and ***electric current*** in the second semester of this course. Physics and math majors will want to consider what happens when they take the quantity v and make it a *vector field* $\vec{v}$ that might *not* be flowing uniformly in the pipe, which might *not* have a uniform shape or cross section, and thence think still more generally to fluids flowing in arbitrary streamlined patterns. Future physicians, however, can draw a graceful curtain across these meditations for the moment, although they too will benefit next semester if they at least *try* to think about them now.

8.4.1: Conservation of Flow

Fluid does not, of course, only flow in smooth pipes with a single cross-sectional area. Sometimes it flows from large pipes into smaller ones or vice versa. We will now proceed to derive an important aspect of that flow for incompressible fluids and/or steady state flows of compressible ones.

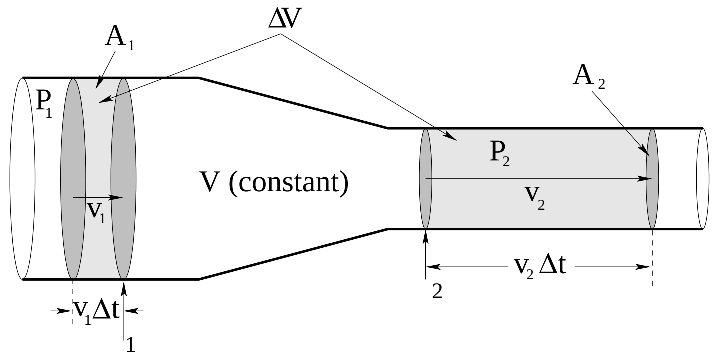


Figure 8.12: Water flows from a wider pipe with a "larger" cros-sectional area A_1 into a narrower pipe with a smaller cross-sectional area A_2. The speed of the fluid in the wider pipe is v_1, in the narrower one it is v_2. The pressure in the wider pipe is P_1, in the narrower one it is P_2.

Figure 8.12 shows a fluid as it flows from just such a wider pipe down a gently sloping neck into a narrower one. As before, we will ignore drag forces and assume that the flow is as uniform as possible as it narrows, while remaining completely uniform in the wider pipe and smaller pipe on either side of the neck. The pressure, speed of the (presumed incompressible) fluid, and cross sectional area for either pipe are P_1, v_1, and A_1 in the wider one and P_2, v_2, and A_2 in the narrower one.

Pay careful attention to the following reasoning. In a time Δt then – as before – a volume of fluid $\Delta V = A_1 v_1 \Delta t$ passes through the surface/past the point 1 marked with an arrow in the figure. In the volume between this surface and the next grey surface at the point 2 marked with an arrow ***no fluid can build up*** so actual quantity of mass in this volume must be a *constant*.

This is very important. The argument is simple. If more fluid flowed into this volume through the first surface than escaped through the second one, then fluid would be *building up* in the volume. This would increase the density. But the fluid's density *cannot* change – it is (by hypothesis) incompressible. Nor can *more* fluid escape through the second surface than enters through the first one.

Note well that this assertion implies that the fluid itself ***cannot be created or destroyed***, it can only flow *into* the volume through one surface and *out* through another, and because it is incompressible and uniform and the walls of the vessel are impermeable (don't leak) the quantity of fluid inside the surface cannot change in any other way.

This is a kind of *conservation law* which, for a continuous fluid or similar medium, is called a *continuity equation*. In particular, we are postulating the law of conservation of matter, implying a continuous flow of matter from one place to another! Strictly speaking, continuity alone would permit fluid to build up in between the surfaces (as this can be managed without creating or destroying the mass of the fluid) but we've trumped that by insisting that the fluid be incompressible.

This means that however much fluid *enters* on the left *must exit on the right* in the time Δt; the shaded volumes on the left and right in the figure above must be *equal*. If we write this out algebraically:

$$\begin{aligned} \Delta V &= A_1 v_1 \, \Delta t = A_2 v_2 \, \Delta t \\ I = \frac{\Delta V}{\Delta t} &= A_1 v_1 = A_2 v_2 \end{aligned} \tag{8.70}$$

Thus the current or flow through the two surfaces marked 1 and 2 must be the *same*:

$$A_1 v_1 = A_2 v_2 \tag{8.71}$$

Obviously, this argument would continue to work if it necked down (or up) further into a pipe with cross sectional area A_3, where it had speed v_3 and pressure P_3, and so on. The flow of water in the pipe must be *uniform*, $I = Av$ must be a *constant independent of where you are in the pipe!*

There are two more meaty results to extract from this picture before we move on, that combine into one "named" phenomenon. The first is that conservation of flow implies that the ***fluid speeds up*** when it flows from a wide tube and into a narrow one or vice versa, it slows down when it enters a wider tube from a narrow one. This means that every little chunk of mass in the fluid on the right is moving *faster* than it is on the left. The fluid has ***accelerated!***

Well, by now you should very well appreciate that *if* the fluid accelerates *then* there must be a net external force that acts on it. The only catch is, where is that force? What exerts it?

The force is exerted by the ***pressure difference*** $\boldsymbol{\Delta P}$ between P_1 and P_2. The force exerted by pressure at the walls of the container points only perpendicular to the pipe at that point; the fluid is moving parallel to the surface of the pipe and hence this "normal" confining force does no work and cannot speed up the fluid.

In a bit we will work out *quantitatively* how much the fluid speeds up, but even now we can see that since $A_1 > A_2$, it must be true that $v_2 > v_1$, and hence $P_1 > P_2$. This is a general result, which we state in words:

The pressure decreases in the direction that fluid velocity increases.

This might well be stated (in other books or in a paper you are reading) the other way: When a fluid slows *down*, the pressure in it *increases*. Either way the result is the same.

This result is responsible for many observable phenomena, notably the mechanism of the lift that supports a frisbee or airplane wing or the ***Magnus effect***[187] that causes a spinning thrown baseball ball to curve.

Unfortunately, treating these phenomena *quantitatively* is beyond, and I do mean *way* beyond, the scope of this course. To correctly deal with lift for compressible or incompressible fluids one must work with and solve either the ***Euler equations***[188] , which are coupled partial

[187] Wikipedia: http://www.wikipedia.org/wiki/Magnus Effect.

[188] Wikipedia: http://www.wikipedia.org/wiki/Euler Equations (fluid dynamics).

differential equations that express Newton's Laws for fluids dynamically moving themselves in terms of the local density, the local pressure, and the local fluid velocity, or the ***Navier-Stokes Equations***[189] , ditto but *including the effects of viscosity* (neglected by Euler). Engineering students (especially those interested in aerospace engineering and real fluid dynamics) and math and physics majors are encouraged to take a peek at these articles, but not too long a peek lest you decide that perhaps majoring in advanced basket weaving really *was* the right choice after all. They are really, really difficult; on the high end "supergenius" difficult[190].

This isn't surprising – the equations have to be able to describe every possible dynamical state of a fluid as it flows in every possible environment – laminar flow, rotational flow, turbulence, drag, around/over smooth shapes, horribly not smooth shapes, and everything in between. At that, they don't account for things like temperature and the mixing of fluids and fluid chemistry – reality is more complex still. That's why we are stopping with the simple rules above – fluid flow is conserved (safe enough) and pressure decreases as fluid velocity increases, all things being equal.

All things are, of course, *not* always equal. In particular, one thing that can easily vary in the case of fluid flowing in pipes is *the height of the pipes*. The increase in velocity caused by a pressure differential can be interpreted or predicted by the Work-Kinetic Energy theorem, but if the fluid is moving *up or down hill* then we may discover that *gravity* is doing work as well!

In this case we should really use the Work-*Mechanical* Energy theorem to determine how pressure changes can move fluids. This is actually pretty easy to do, so let's do it.

8.4.2: Work-Mechanical Energy in Fluids: Bernoulli's Equation

Daniel Bernoulli was a third generation member of the famous Bernoulli family[191] who worked on (among many other things) fluid dynamics, along with his good friend and contemporary, Leonhard Euler. In 1738 he published a long work wherein he developed the idea of energy conservation to fluid motion. We'll try to manage it in a page or so.

In figure 8.13 we see the same pipe we used to discuss conservation of flow, only now it is bent uphill so the 1 and 2 sections of the pipe are at heights y_1 and y_2 respectively. This really is the only change, otherwise we will recapitulate the same reasoning. The fluid is incompressible and the pipe itself does not leak, so fluid cannot build up between the bottom and the top. As the fluid on the bottom moves to the left a distance d (which might be $v_1\Delta t$ but we don't insist on it as rates will not be important in our result) exactly the same amount fluid must move to the left a distance D up at the top so that fluid is conserved.

The total *mechanical* consequence of this movement is thus the disappearance of a chunk of fluid mass:

$$\Delta m = \rho\Delta V = \rho A_1 d = \rho A_2 D \tag{8.72}$$

[189] Wikipedia: http://www.wikipedia.org/wiki/Navier-Stokes equations.

[190] To give you an idea of *how* difficult they are, note that there is a $1,000,000 prize just for showing that solutions to the 3 dimensional Navier-Stokes equations generally *exist* and/or are *not singular*.

[191] Wikipedia: http://www.wikipedia.org/wiki/Bernoulli family. The Bernoullis were in on many of major mathematical and physical discoveries of the eighteenth and nineteenth century. Calculus, number theory, polynomial algebra, probability and statistics, fluid dynamics – if a theorem, distribution, principle has the name "Bernoulli" on it, it's gotta be good...

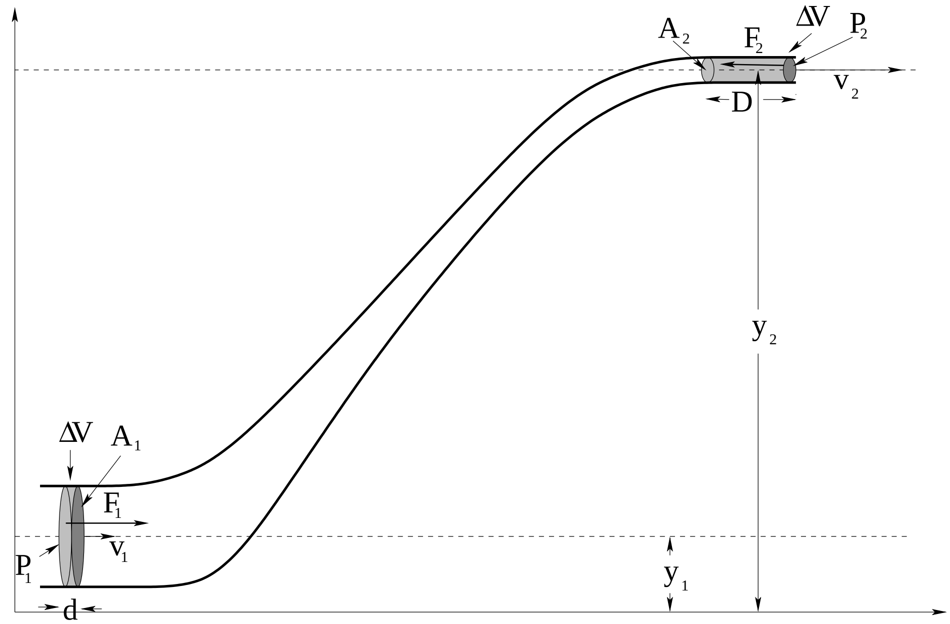


Figure 8.13: A circular cross-sectional necked pipe is arranged so that the pipe *changes height* between the larger and smaller sections. We will assume that both pipe segments are narrow compared to the height change, so that we don't have to account for a potential energy difference (per unit volume) between water flowing at the top of a pipe compared to the bottom, but for ease of viewing we do not draw the *picture* that way.

that is moving at speed v_1 and at height y_1 at the bottom and it's appearance moving at speed v_2 and at height y_2 at the top. Clearly ***both*** the kinetic energy ***and*** the potential energy of this chunk of mass have changed.

What caused this change in mechanical energy? Well, it can only be *work*. What does the work? The walls of the (frictionless, drag free) pipe can do no work as the only force it exerts is perpendicular to the wall and hence to $\vec{v}$ in the fluid. The only thing left is the ***pressure*** that acts on the entire block of water between the first surface (lightly shaded) drawn at both the top and the bottom as it moves forward to become the second surface (darkly shaded) drawn at the top and the bottom, effecting this net transfer of mass Δm.

The force F_1 exerted to the *right* on this block of fluid at the bottom is just $F_1 = P_1 A_1$; the force F_2 exerted to the *left* on this block of fluid at the top is similarly $F_2 = P_2 A_2$. The work done by the pressure acting over a distance d at the bottom is $W_1 = P_1 A_1 d$, at the top it is $W_2 = -P_2 A_2 D$. The total work is equal to the total change in mechanical energy of the chunk

Δm:

$$
\begin{aligned}
W_{\text{tot}} &= \Delta E_{\text{mech}} \\
W_1 + W_2 &= E_{\text{mech}}(final) - E_{\text{mech}}(initial) \\
P_1 A_1 d - P_2 A_2 D &= (\frac{1}{2}\Delta m v_2^2 + \Delta m g y_2) - (\frac{1}{2}\Delta m v_1^2 + \Delta m g y_1) \\
(P_1 - P_2)\Delta V &= (\frac{1}{2}\rho \Delta V v_2^2 + \rho \Delta V g y_2) - (\frac{1}{2}\rho \Delta V v_1^2 + \rho \Delta V g y_1) \\
(P_1 - P_2) &= (\frac{1}{2}\rho v_2^2 + \rho g y_2) - (\frac{1}{2}\rho v_1^2 + \rho g y_1) \\
P_1 + \frac{1}{2}\rho v_1^2 + \rho y_1 &= P_2 + \frac{1}{2}\rho v_2^2 + \rho g y_2 \ = \ \text{a constant (units of pressure)} \qquad (8.73)
\end{aligned}
$$

There, that wasn't so difficult, was it? This lovely result is known as ***Bernoulli's Principle*** (or the Bernoulli fluid equation). It contains pretty much *everything* we've done so far except conservation of flow (which is a distinct result, for all that we used it in the derivation) and Archimedes' Principle.

For example, if $v_1 = v_2 = 0$, it describes a static fluid:

$$P_2 - P_1 = -\rho g(y_2 - y_1) \qquad (8.74)$$

and if we change variables to make z (depth) $-y$ (negative height) we get the familiar:

$$\Delta P = \rho g \Delta z \qquad (8.75)$$

for a static incompressible fluid. It also not only tells us that pressure drops where fluid velocity increases, it tells us *how much* the pressure drops when it increases, allowing for things like the fluid flowing up or downhill at the same time! Very powerful formula, and all it is is the Work-Mechanical Energy theorem (per unit volume, as we divided out ΔV in the derivation, note well) applied to the fluid!

Example 8.4.1: Emptying the Iced Tea

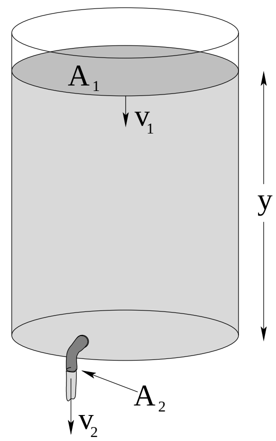


Figure 8.14

In figure 8.14 above, a cooler full of iced tea is displayed. A tap in the bottom is opened, and the iced tea is running out. The cross-sectional area of the top surface of the tea (open to the atmosphere) is A_1. The cross-sectional area of the tap is $A_2 \ll A_1$, and it is also open to the atmosphere. The depth of iced tea (above the tap at the bottom) is y. The density of iced tea is basically identical to that of water, ρ_w. Ignore viscosity and resistance and drag.

What is the speed v_2 of the iced tea as it exits the tap at the bottom? How rapidly is the top of the iced tea descending at the top v_1? What is the rate of flow (volume per unit time) of the iced tea when the height is y?

This problem is fairly typical of Bernoulli's equation problems. The two concepts we will need are:

a) Conservation of flow: $A_1 v_1 = A_2 v_2 = I$

b) Bernoulli's formula: $P + \rho g y + \frac{1}{2}\rho v^2 = \mathrm{constant}$

First, let's write Bernoulli's formula for the top of the fluid and the fluid where it exits the tap. We'll choose $y = 0$ at the height of the tap.

$$P_0 + \rho g y + \frac{1}{2}\rho v_1^2 = P_0 + \rho g(0) + \frac{1}{2}\rho v_2^2 \tag{8.76}$$

We have two unknowns, v_1 and v_2. Let's eliminate v_1 in favor of v_2 using the flow equation and substitute it into Bernoulli.

$$v_1 = \frac{A_2}{A_1} v_2 \tag{8.77}$$

so (rearranging):

$$\rho g y = \frac{1}{2}\rho v_2^2 \left\{ 1 - \left(\frac{A_2}{A_1}\right)^2 \right\} \tag{8.78}$$

At this point, since $A_1 \gg A_2$ we will often want to approximate:

$$\left(\frac{A_2}{A_1}\right)^2 \approx 0 \tag{8.79}$$

and solve for

$$v_2 \approx \sqrt{2gy} \tag{8.80}$$

but it isn't *that* difficult to leave the factor in.

This latter result (equation 8.80) is known as *Torricelli's Law*. Torricelli is known to us as the inventor of the barometer, as a moderately famous mathematician, and as the final secretary of Galileo Galilei in the months before his death in 1642 – he completed the last of Galileo's dialogues under Galileo's personal direction and later saw to its publication. It basically states that a fluid exits any (sufficiently large) container through a (sufficiently small) hole at the same speed a rock would have if dropped from the height from the top of the fluid to the hole. This was a profound observation of gravitational energy conservation in a context quite different from Galileo's original observations of the universality of gravitation.

Given this result, it is now trivial to obtain v_1 from the relation above (which should be quite accurate even if you assumed the ratio to be zero initially) and compute the rate of flow from e.g. $I = A_2 v_2$. An interesting exercise in calculus is to estimate the time required for the vat of iced tea to empty through the lower tap if it starts at initial height y_0. It requires realizing that $\frac{dy}{dt} = -v_1$ and transforming the result above into a differential equation. Give it a try!

Example 8.4.2: Flow Between Two Tanks

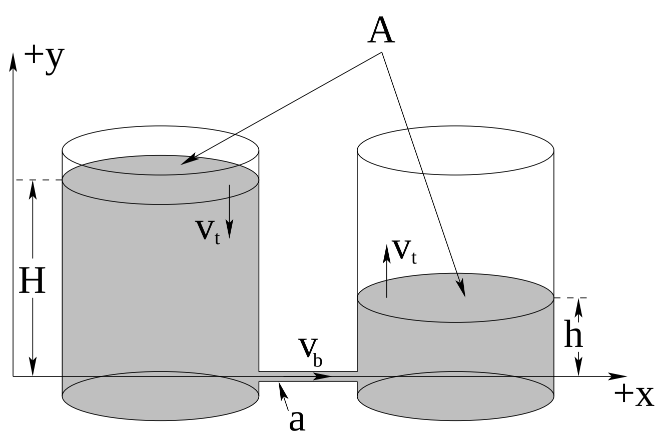


Figure 8.15

In figure 8.15 two water tanks are filled to different heights. The two tanks are connected at the bottom by a narrow pipe through which water (density ρ_w) flows *without resistance* (see the next section to understand what one might do to include resistance in this pipe, but for the moment this will be considered too difficult to include in an introductory course). Both tanks are open to ordinary air pressure P_0 (one atmosphere) at the top. The cross sectional area of both tanks is A and the cross sectional area of the pipe is $a \ll A$.

Once again we would like to know things like: What is the speed v_b with which water flows through the small pipe from the tank on the left to the tank on the right? How fast does the water level of the tank on the left/right fall or rise? How long would it take for the two levels to become equal, starting from heights H and h as shown?

As before, we will need to use:

a) Conservation of flow: $A_1 v_1 = A_2 v_2 = I$

b) Bernoulli's formula: $P + \rho g y + \frac{1}{2}\rho v^2 = \text{constant}$

This problem is actually more difficult than the previous problem. If one naively tries to express Bernoulli for the top of the tank on the left and the top of the tank on the right, one gets:

$$P_0 + \rho g H + \frac{1}{2}\rho v_t^2 = P_0 + \rho g h + \frac{1}{2}\rho v_t^2 \tag{8.81}$$

Note that we've equated the speeds *and* pressures on both sides because the tanks have equal cross-sectional areas so that they have to be the same. But this makes no sense! $\rho g H \neq \rho g h$!

The problem is this. The pressure is not constant across the pipe on the bottom. If you think about it, the pressure at the bottom of a nearly static fluid column on the left (just outside the mouth of the pipe) has to be approximately $P_0 + \rho g H$ (we can and will do a bit better than this, but this is what we expect when $a \ll A$). The pressure just outside of the mouth of the pipe in the fluid column on the right must be $P_0 + \rho g h$ from the same argument. *Physically*, it is this pressure difference that forces the fluid through the pipe, speeding it up as it enters on the left. The pressure in the pipe has to *drop* as the fluid speeds up entering the pipe from the bottom of the tank on the left (the Venturi effect).

This suggests that we write Bernoulli's formula for the top of the left hand tank and a point just inside the pipe at the bottom of the left hand tank:

$$P_0 + \rho g H + \frac{1}{2}\rho v_t^2 = P_p + \rho g(0) + \frac{1}{2}\rho v_b^2 \tag{8.82}$$

This equation has three unknowns: v_t, v_b, and P_p, the pressure just inside the pipe. As before, we can easily eliminate e.g. v_t in favor of v_b:

$$v_t = \frac{a}{A} v_b \tag{8.83}$$

so (rearranging):

$$P_0 - P_p + \rho g H = \frac{1}{2}\rho v_b^2 \left\{1 - \left(\frac{a}{A}\right)^2\right\} \tag{8.84}$$

Just for fun, this time we won't approximate and throw away the a/A term, although in most cases we could and we'd never be able to detect the perhaps 1% or even less difference.

The problem now is: What is P_p? That we get on the other side, but not the way you might expect. Note that the pressure must be *constant* all the way across the pipe as neither y nor v can change. The pressure in the pipe must therefore match the pressure at the bottom of the other tank. But that pressure is just $P_0 + \rho g h$! Note well that this is completely consistent with what we did for the iced tea – the pressure at the outflow had to match the air pressure in the room.

Substituting, we get:

$$\rho g H - \rho g h = \frac{1}{2}\rho v_b^2 \left\{1 - \left(\frac{a}{A}\right)^2\right\} \tag{8.85}$$

or

$$v_b = \sqrt{\left\{ \frac{2g(H-h)}{1-\left(\frac{a}{A}\right)^2} \right\}} \tag{8.86}$$

which ***makes sense!*** What pushes the fluid through the pipe, speeding it up along the way? The difference in pressure between the ends. But the pressure inside the pipe itself has to match the pressure at the outflow because it has *already accelerated to this speed* across the tiny distance between a point on the bottom of the left tank "outside" of the pipe and a point just inside the pipe.

From v_b one can as before find v_t, and from v_t one can do some calculus to at the very least write an integral expression that can yield the time required for the two heights to come to equilibrium, which will happen when $H - \int v_t(t)\,dt = h + \int v_t(t)\,dt$. That is, the rate of change of the *difference* of the two heights is twice the velocity of either side.

Note well: This example is also highly relevant to the way a ***siphon*** works, whether the siphon empties into air or into fluid in a catch vessel. In particular, the pressure drops at the intake of the siphon to match the height-adjusted pressure at the siphon outflow, which could be in air or fluid in the catch vessel. The main difference is that the siphon tube is *not horizontal*, so according to Bernoulli the pressure has to increase or decrease as the fluid goes up and down in the siphon tube at constant velocity.

Questions: Is there a maximum height a siphon can function? Is the maximum height measured/determined by the intake side or the outflow side?

8.4.3: Fluid Viscosity and Resistance

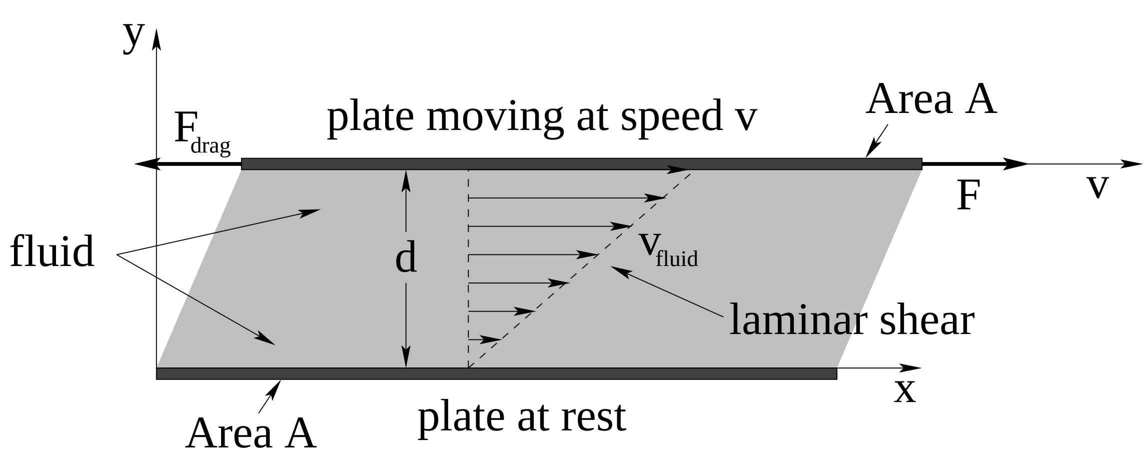


Figure 8.16: Dynamic viscosity is defined by the scaling of the force F required to keep a plate of cross-sectional area A moving at constant speed v on top of a layer of fluid of thickness d. This causes the fluid to ***shear***. Shear stress is explained below and in more detail in the chapter on oscillations.

In the discussion above, we have consistently ignored viscosity and drag, which behave like "friction", exerting a force *parallel* to the confining walls of a pipe in the direction opposite to the relative motion of fluid and pipe. We will now present a very simple (indeed, oversimplified!) treatment of viscosity, but one that (like our similarly oversimplified treatment of static and kinetic friction) is sufficient to give us a good conceptual understanding of what viscosity is and that will work well quantitatively – until it doesn't.

In figure 8.16 above a two horizontal plates with cross-sectional areas A have a layer of fluid with thickness d trapped between them. The plates are assumed to be very large, with a comparatively thin layer of fluid in between, so that we can neglect what happens near their edges. The bottom plate is fixed in our (inertial) frame of reference and the upper plate is moving to the right, pushed by a *constant* force F so that it maintains a ***constant speed*** $\boldsymbol{v}$ against the drag force exerted on it by the viscous fluid.

Many fluids, especially "wet" liquids like water (a polar molecule) strongly interact with the solid wall and basically stick to it at very short range – this is known as the "no-slip condition". This means that a thin layer (a few molecules thick) of the fluid in "direct contact" with the upper or lower plates will generally not be moving *relative* to the plate it is touching. That is, in the figure the fluid is at rest where it touches the bottom plate, and is moving at speed v (in the x-direction) where it touches the top plate. In between the speed of the fluid must vary from 0 at the bottom to v at the top.

When the speed v is not too large (and various other parametric stars are in alignment), many fluids can be treated as ***Newtonian fluids***: layers of fluid such that each layer is going a bit faster than the layer below it and a bit slower than the one above it all the way from the bottom fixed layer all the way up to the top layer stuck to the moving plate. As discussed in week 2, this is called ***laminar flow***, and results in this simple case in the *linear* velocity profile illustrated in 8.16 above.

Consider two adjacent layers in the fluid with an imaginary surface separating them. The lower layer is moving more slowly than the upper one and hence there is "dynamic friction" between the two layers across the surface. This friction pulls *forward* (in the direction of the movement of the top plate) on the lower layer and *backward* on the upper layer. For each differentially thin layer in between the plates, then, there is a force pulling it forward from the layer above it and a force pulling it backwards from the layer below it, and in dynamic equilibrium (where the layers are all moving at their own constant speeds) these forces have to be *equal and opposite*.

We see that the total force pulling each layer *forward* has to *strictly increase* as one moves from the bottom plate/layer to the top plate/layer. To keep the top plate moving at a constant speed, it has to be pushed *forward* by an external force equal to the ***total (shear) force accumulated across all the layers*** that makes up the drag force acting opposite to its velocity relative to the bottom plate.

For a *Newtonian fluid*[192] it is empirically observed that the (shear) force required to keep the top plate moving at speed v in the x direction relative to the fixed bottom plate in figure 8.16 is:

$$F = \mu A \frac{v}{d} \tag{8.87}$$

This formula incidentally ***defines*** the constant of proportionality μ, characteristic of the

[192]Wikipedia: http://www.wikipedia.org/wiki/Newtonian Fluid. More advanced students will want to note that the formula presented here is just a part of a much more general partial differential equation for the *shear stress tensor*, and that the math required to do fluid dynamics correctly is not at all as simple as it might appear from these idealized cases.

fluid, called the ***dynamic viscosity***. It is the moral equivalent of the "coefficient of kinetic friction" between fluid layers. However, if we divide the force F by the area A in contact we get the ***shear stress*** F/A and that this has SI *units of pressure*, pascals, but is ***not*** a pressure. The shear force F is directed *parallel* to the surface area A and not perpendicular to it. The units of v/d are inverse time. Consequently, the SI units of μ are pascal-seconds, unlike the coefficients of friction which are dimensionless.

This expression gives the *total* force transmitted from the bottom plate to the top plate through all of the layers of the fluid in between. Since the geometry of the figure is very simple, we expect the force to be distributed more or less *linearly* across the many layers in between. We express this assumption by looking at two layers of fluid, one below and one above, as drawn in figure 8.17.

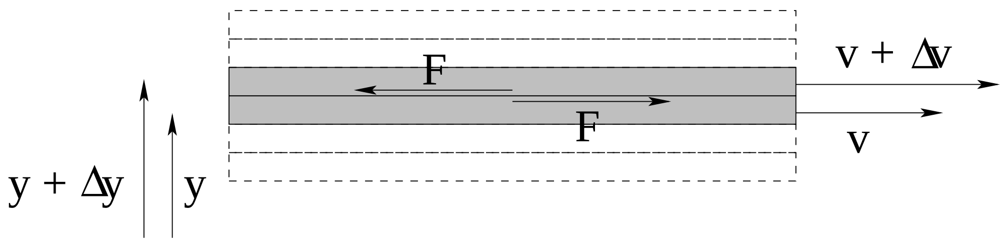


Figure 8.17: Two layers of fluid moving at slightly different velocities exert an equal and opposite force on each other at the boundary in between.

In order to get the empirical result above, we expect that the magnitude of the force between the layers is given by:

$$F = \mu A \frac{\Delta v}{\Delta y} \tag{8.88}$$

Note well that this is not the *net* force on a layer! The lower (shaded) layer in the figure above acts to slow the (shaded) layer just above, dragging it backwards, as the upper layer acts to drage the lower forward. But the layer *above* the upper layer *also* acts on the upper layer with the exact same magnitude of force, dragging it forward. The total *vector* force on a layer is thus *zero*, as must be the case for the layers in the fluid to have constant speed in dynamic equilibrium!

We would like to be able to use this expression for fluid layers that are not simple plane sheets between two plates, where we can't guess that the velocity profile of the fluid is a simple linear variation from one boundary surface to another. In that case, we expect that – subject to various conditions, such as the areas in contact being at least *locally* flat – the magnitude of the force between neighboring layers is proportional to the (partial) *derivative* of the fluid velocity in the direction of flow with respect to the direction perpendicular to the flow:

$$F = \mu A \frac{\partial v}{\partial y} \tag{8.89}$$

This expression isn't quite correct – it is a single part of a more general tensor form – but it will do for the next section, where we give a simple derivation of the velocity profile and drag resistance of a circular pipe.

8.4.4: Resistance of a Circular Pipe: Poiseuille's Equation

The case of greatest interest to many of the students taking this course is not, actually, the drag force between parallel plates separated by a viscous fluid. It is the case of a viscous fluid in laminar flow in a ***pipe with a circular cross-section***. This is an excellent model for water flowing in ordinary pipes as well as blood flowing in arteries and veins in living organisms. It is thus useful to engineers and life science students alike, and of course physics or math majors should become as familiar as possible with *all* simple systems of fluid dynamics in preparation to tackling (eventually) the fiendish difficulty of the Navier-Stokes equation.

Note Well!

This section is split at the line below (as are several others in this textbook) into two parts. In the first, we explicitly step through at least one (comparatively simple, believe it or not) derivation of the Hagen-Poiseuille equation[193]. After the *second* separator line we will start *with* the Poiseuille equation as a given result and observe its general scaling and how to apply it without worrying about exactly where it comes from.

Physics and math majors (and perhaps some engineering students interested in fluid engineering, aerodynamics, and so on) *should* at least read through and study the actual derivation, and *all* students are welcome to take a look, but *all* students (even engineers and physics majors) can skip all of the content between the two separator lines and start reading again with the conceptual/application part that follows and still do perfectly well on the actual problems.

As was the case in our discussion of shear stress and viscosity above, we expect a viscous fluid moving at a reasonably low speed in a circular pipe to stick to the walls, creating a boundary layer of fluid that does not move. We will therefore make this "no-slip" condition a boundary condition of our problem. Unlike the shear stress example above, however, there is no possibility of moving two surfaces to maintain a laminar velocity profile in between, as there is only *one* surface bounding the fluid flowing in a pipe, at rest!

Let's assume that the fluid is moving slowly enough that laminar flow is established inside the pipe in steady state. The fluid touching the walls is not moving from the no-slip condition, so there must be an *annular velocity profile* where each annular layer is moving a bit faster as one moves in from the edges from the pipe to the center. Figure **??** can help you visualize this sort of laminar flow where the highest speed of the fluid is found in the center of the pipe, the greatest possible distance from the walls.

Here we cannot assume that the velocity profile is linear with the distance from the pipe walls, and in fact it turns out *not* to be linear. We have to apply the force rule we deduced for shear forces between layers in the previous section to try to obtain a differential equation representing Newton's Second Law for an *arbitrary* thin annular layer of fluid at radius r and with thickness dr. This is just *one* of the many annular cylindrical shell's of fluid portrayed above and is illustrated in figure 8.19.

[193] For the record, Poiseuille is correctly pronounced "Pwa-zoo-ee" – pwazooee. Not "poise-you-ill" or any of the myriad of other ways this might be interpreted by English-speaking students. And no, I didn't get it right either and had to use the Internet to look it up...

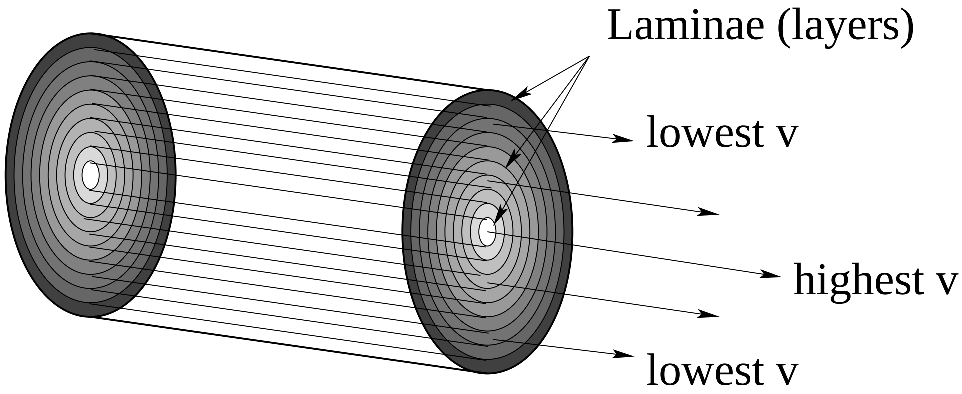


Figure 8.18: Cross-section of laminar flow in a pipe, where the darker shades correspond to slower-moving fluid in concentric layers (laminae) from the wall of the pipe (slowest speed) to the center (highest speed).

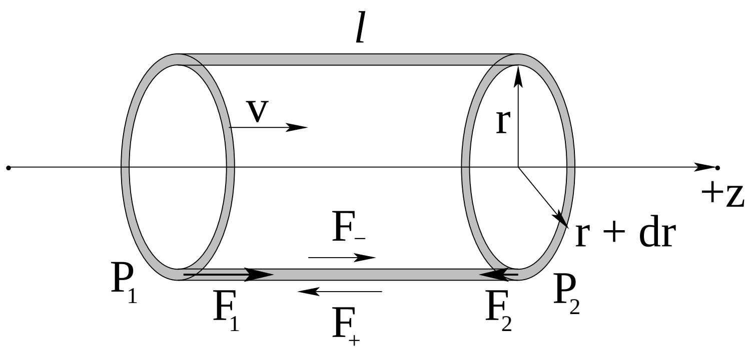


Figure 8.19: A differentially thin annular cylindrical shell of the fluid in the pipe. The pressure difference across the ends of the shell exerts a net force to the right. The differential viscous force between the inside and out surfaces of the shell exerts a net drag force on the shell. These must balance, maintaining a constant speed for the laminar shell v.

In order for this laminar shell to move through the pipe in the direction indicated with a contant speed v (which so far is an unknown function of radius r) the forces on it must balance. There are two distinct sources of force that push on the shell. One is there must be a *pressure difference* between the pressure P_1 at the left side of the shell and P_2 at the right side of the shell, with $P_1 > P_2$, pushing the shell to the right. The other is the drag force due to viscosity acting on both the inner (-) and outer (+) surfaces of the layer, at the radii r and $r + dr$ respectively. We expect the force on the *inner* layer to be *in* the direction of fluid flow as the fluid is moving faster than the fluid in the shell there, and the force on the outer layer to be *against* the direction of fluid flow as the fluid outside of the shell is moving more slowly than the fluid in the shell.

We will use the differential form of the formula we deduced in the previous section to describe the drag forces between the surfaces of the shell and the surrounding fluid. The area of the inside of the annular cylinder in contact with the faster moving fluid towards the center is $2\pi r\ell$. We therefore expect the drag force due to viscosity acting on this layer to be:

$$F_{-} = -\mu(2\pi r\ell)\left.\frac{dv}{dr}\right|_{r} \tag{8.90}$$

Note that with the minus sign this force points towards the *right* – in the direction of flow – as expected because dv/dr is *negative*, with the highest speed at $r = 0$ in the middle.

The layer just *outside* of the illustrated laminar shell also exerts a drag force on our annular cylinder. The fluid there is moving more slowly, so it *slows it down*. We can use exactly the same relation to determine this force, but the area in contact is slightly differentially larger and the derivative has to be evaluated at a slightly (differentially) larger value of r:

$$F_+ = \mu(2\pi(r+dr)\ell)\left.\frac{dv}{dr}\right|_{r+dr} \tag{8.91}$$

Again, dv/dr is negative, so this (z-directed) force points to the left as expected.

Finally, there are the two forces $F_1 = P_1 dA$ and $F_2 = P_2 dA$ acting on the ends of the shell due to the pressure. The cross-sectional area of the cylindrical shell is $dA = 2\pi r dr$ (its circumference times its thickness) so:

$$F_{\Delta P} = F_1 - F_2 = (P_1 - P_2)dA = \Delta P(2\pi r dr) \tag{8.92}$$

Assembling the entire expression, force balance is thus:

$$\Delta P(2\pi r dr) - \mu(2\pi r\ell)\left.\frac{dv}{dr}\right|_r + \mu(2\pi(r+dr)\ell)\left.\frac{dv}{dr}\right|_{r+dr} = 0 \tag{8.93}$$

So far, this has been pretty straightforward. Now things get a bit tricky. We use a Taylor series expansion for the derivative in the last term so we can evaluate all the derivatives in question at r only:

$$\left.\frac{dv}{dr}\right|_{r+dr} = \left.\frac{dv}{dr}\right|_r + \left.\frac{d^2v}{dr^2}\right|_r dr \tag{8.94}$$

or:

$$\Delta P(2\pi r dr) - \mu(2\pi r\ell)\frac{dv}{dr} + \mu(2\pi(r+dr)\ell)\left\{\frac{dv}{dr} + \frac{d^2v}{dr^2}dr\right\} = 0 \tag{8.95}$$

where all derivatives are evaluated at r so we no longer need to indicate the limits explicitly.

Two terms cancel, and we end up with:

$$\Delta P(2\pi r dr) + \mu(2\pi\ell)\frac{dv}{dr}dr + \mu(2\pi r\ell)\frac{d^2v}{dr^2}dr + \mu(2\pi\ell)\frac{d^2v}{dr^2}(dr)^2 = 0 \tag{8.96}$$

We then divide out the common factors of $2\pi r dr$ and throw out the last term as it vanishes like dr. Finally we rearrange a bit to get:

$$-\frac{1}{\mu}\frac{\Delta P}{\ell} = \left\{\frac{1}{r}\frac{dv}{dr} + \frac{d^2v}{dr^2}\right\} = \frac{1}{r}\frac{d}{dr}\left(r\frac{dv}{dr}\right) \tag{8.97}$$

The right hand side is (as it turns out) the radial part of the Laplacian in cylindrical coordinates, so this is a form of the Poisson (inhomogeneous Laplace) equation. Integrating this is beyond the scope of this course but is straightforward and taught in more advanced math and physics classes. We will simply jump to the result. If we integrate this differential equation subject to the boundary conditions $v(R) = 0$ for a pipe of radius R, we get:

$$v(r) = \frac{1}{4\mu}\frac{\Delta P}{\ell}\left(R^2 - r^2\right) \tag{8.98}$$

We can write this one last way:

$$v(r) = (1/4\mu)R^2\frac{\Delta P}{\ell}\left\{1 - \left(\frac{r}{R}\right)^2\right\} = v_{\max}\left\{1 - \left(\frac{r}{R}\right)^2\right\} \tag{8.99}$$

where $v_{\max}$ is the maximum speed of the fluid in the center of the pipe.

This is a *quadratic* profile to the velocity cross-section, peaked in the center, quite different from the linear profile we expected from symmetry if nothing else for two flat plates.

We are finally in position to get Poiseuille's Equation and determine the ***resistance*** to flow of a circular pipe carrying a fluid as a function of its parameters. Note well that the volume of fluid being carried from left to right per unit time – the *total volumetric current* dI or the *volumetric flow* of the annular shell – is just this speed times the cross-sectional area:

$$dI = \frac{dV}{dt} = vdA = (1/4\mu)R^2\frac{\Delta P}{\ell}\left\{1 - \left(\frac{r}{R}\right)^2\right\}2\pi r dr \tag{8.100}$$

We can easily integrate this from 0 to R to find the total current:

$$I = \int dI = \frac{\pi R^2}{2\mu}\frac{\Delta P}{\ell}\int_0^R \left(r - \frac{r^3}{R^2}\right)dr = \frac{\pi R^2}{2\mu}\frac{\Delta P}{\ell}\frac{R^2}{4} = \frac{\pi R^4}{8\mu}\frac{\Delta P}{\ell} \tag{8.101}$$

This equation relates the volumetric current in a circular pipe to the pressure difference one must maintain across the pipe in order to overcome the drag force between the walls of the pipe and the moving fluid. This concludes this section.

In the previous, possibly omitted, section we saw that – after a great deal of work – we could show that for a *circular pipe of radius r and length ℓ, carrying a fluid with dynamic viscosity μ in laminar flow*:

$$\Delta P = I\frac{8\mu\ell}{\pi r^4} \tag{8.102}$$

where I is the volumetric current, ΔP is the pressure difference across the pipe. Note well that we switched variables from R being the radius of the pipe to r being the radius in order to not confuse it with the pipe's resistance below. This is an example of a more general relation that holds for *any* pipe with a uniform, given, cross-section:

$$\Delta P = IR \tag{8.103}$$

where R is called the ***resistance*** of the pipe. Obviously,

$$R = \frac{8\mu\ell}{\pi r^4} \tag{8.104}$$

for a normal circular pipe (or vein, or artery, or hydraulic fluid line, or...).

This expression is the analog, in fluid mechanics, of *Ohm's Law* in electrical circuits. The pressure difference across the pipe is directly proportional to the rate of fluid flow. Double the pressure difference and you get twice as much fluid out of a given pipe (all things being equal, especially the maintenance of laminar flow). However, for a *fixed* pressure difference, you can also get more or less water out of a pipe by varying its parameters and hence its resistance.

If you double the length of a pipe, this distributes the pressure difference over twice as much pipe so you expect to get half of the flow. If you increase the viscosity – the "stickiness" of the fluid – you expect to get less fluid through the pipe for a given pressure difference. Finally, if you reduce the cross-sectional area of a pipe, you expect to get less fluid through it for a given length, viscosity, and pressure difference.

The only real *surprise* in this is that the resistance goes down like the area of the pipe *squared* (like the fourth power of r). This is really quite unexpected, and arises because the velocity of the fluid in the pipe is not uniform over the cross-sectional area. When you double r you do indeed double the area (and so expect on that basis alone a doubling of the volume), but you also double the maximum velocity of the fluid in the pipe, which results in a *second* doubling!

Equation 8.102 is known as ***Poiseuille's Law*** and is a key relation for physicians, plumbers, physicists, and engineers to know because it describes both *flow of water in pipes* and the *flow of blood in blood vessels* whereever the flow is slow enough that it is laminar and not turbulent (which is actually "mostly", so that the expression is *useful*).

8.4.5: Turbulence

Although turbulence has been *observed* by humans from prehistoric times to the present, the systematic observation and description of turbulence begins with none other than Leonardo da Vinci. Da Vinci observed and sketched the turbulent flow that results when a stream of water falls into a pool, and noted that in the chaos where the two met there were eddies[194] of all sizes, from large ones down to tiny microscopic ones. He noted that large objects responded only to the large rotations, but that tiny objects were swept along and rotated by eddies of all sizes, eddies within eddies within eddies. Da Vinci named this phenomena "turbolenza" – hence our modern term for it.

We have already seen (in week/chapter 2) that there is a dramatic shift in the drag force exerted on an object by a turbulent fluid compared to that of a fluid in laminar flow. In equation 8.87 above, we obtained a justification of sorts for a linear dependence of drag on relative fluid velocity as long as the velocity profile in the fluid near the surface was approximately linear. Of course for fluid flowing in a circular pipe this profile was *not* linear, but if we look more carefully at the form of the force between layers adjacent layers, as long as dy is perpendicular to dv_x:

$$dF = \mu A \frac{dv_x}{dy} dy = \mu A dv_x \tag{8.105}$$

You will not be responsible for the formulas or numbers in this section, but you should be conceptually aware of the *phenomenon* of the "onset of turbulence" in fluid flow. If we return to our original picture of laminar flow between two plates above, consider a small chunk of fluid somewhere in the middle. We recall that there is a (small) shear force across this constant-velocity chunk, which means that there is a drag force to the right at the top and to the left at the bottom. These forces form a ***force couple*** (two equal and opposite forces that do not act along the same line) which exerts a ***net torque on the fluid block***.

[194] An "eddy" is also known as a "whirlpool" or "vortex" – a volume of water that can be almost any size that is *rotating* around a central line of some finite length that may not be straight.

Here's the pretty problem. If the torque is *small* (whatever that word might mean relative to the parameters such as density and viscosity that describe the fluid) then the fluid will ***deform continually*** as described by the laminar viscosity equations above, without actually rotating. The fluid will, in other words, shear instead of rotate in response to the torque. However, as one increases the shear (and hence the velocity gradient and peak velocity) one can get to where the torque across some small cube of fluid causes it to rotate faster than it can shear! Suddenly small *vortices* of fluid appear throughout the laminae! These tiny whirling tubes of fluid have axes (generally) perpendicular to the direction of flow and *add a chaotic, constantly changing structure to the fluid*.

Once tiny vortices start to appear and reach a certain size, they rapidly grow with the velocity gradient and cause a change in the *character* of the drag force coupling across the fluid. There is a dimensionless scale factor called the ***Reynolds Number*** Re[195] that has a certain characteristic value (different for different pipe or plate geometries) at the point where the vortices start to grow so that the fluid flow becomes ***turbulent*** instead of ***laminar***.

The Reynolds number for a circular pipe is:

$$Re = \frac{\rho v D}{\mu} = \frac{\rho v \; 2r}{\mu} \tag{8.106}$$

where $D = 2r$ is the ***hydraulic diameter***[196] , which in the case of a circular pipe is the actual diameter. If we multiply this by one in a particular form:

$$Re = \frac{\rho v D}{\mu}\frac{Dv}{Dv} = \frac{\rho v^2 D^2}{\mu D v} \tag{8.107}$$

$\rho v^2 D^2$ has units of momentum per unit time (work it out). It is proportional to the *inertial force* acting on a differential slice of the fluid. $\mu D v$ *also* has units of force (recall that the units of μ are N-sec/m^2). This is proportional to the "viscous force" propagated through the fluid between layers. The Reynolds number can be thought of as a ratio between the force pushing the fluid forward and the shear force holding the fluid together against rotation.

The one thing the Reynolds number does for *us* is that it serves as a ***marker for the transition to turbulent flow***. For $Re < 2300$ flow in a circular pipe is *laminar* and all of the relations above hold. Turbulent flow occurs for $Re > 4000$. In between is the region known as the ***onset of turbulence***, where the ***resistance of the pipe depends on flow in a very nonlinear fashion***, and among other things *dramatically increases* with the Reynolds number to eventually be proportional to v. As we will see in a moment (in an example) this means that the partial occlusion of blood vessels can have a profound effect on the human circulatory system – basically, instead of $\Delta P = IR$ one has $\Delta P = I^2 R'$ so doubling flow at constant resistance (which may itsef change form) requires four times the pressure difference in the turbulent regime!

At this point you should understand fluid statics and dynamics quite well, armed both with equations such as the Bernoulli Equation that describe idealized fluid dynamics and statics as well as with conceptual (but possibly quantitative) ideas such Pascal's Principle or Archimedes' principle as the relationship between pressure differences, flow, and the geometric factors that contribute to resistance. Let's put some of this nascent understanding to the test by looking over and analyzing the human circulatory system.

[195]Wikipedia: http://www.wikipedia.org/wiki/Reynolds Number.
[196]Wikipedia: http://www.wikipedia.org/wiki/hydraulic diameter.

8.5: The Human Circulatory System

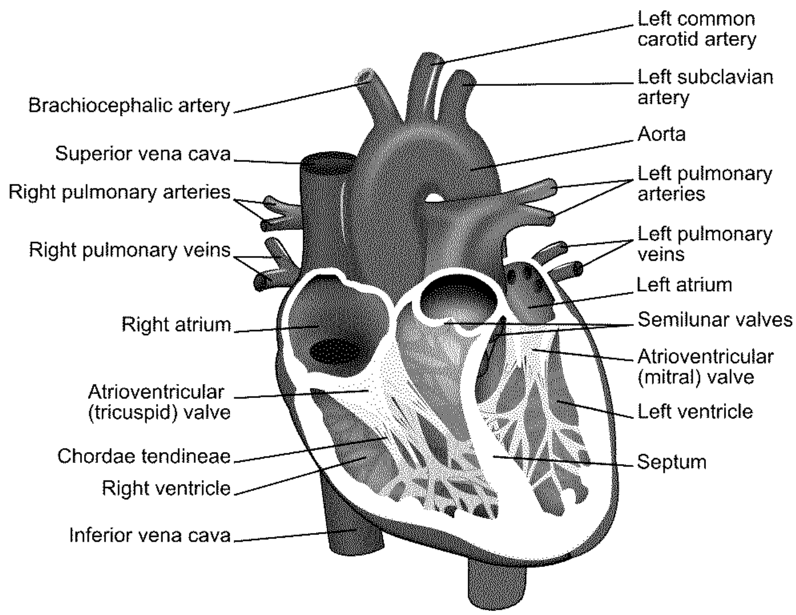


Figure 8.20: A simple cut-away diagram of the human heart.

For once, this is a chapter that math majors, physics majors, and engineers may, if they wish, skip, although personally I think that any intellectually curious person would want to learn all sorts of things that *sooner or later will impact on their own health and life*. To put that rather bluntly, kids, sure ***now*** you're all young and healthy and everything, but in thirty or forty more years (if you survive) you *won't be*, and understanding the things taught in this chapter will be extremely useful to you then, if not now as you choose a lifestyle and diet that might get you *through* to then in reasonably good cardiovascular shape!

Here is a list of True Facts about the human cardiovascular system, in no particular order, that you should now be able to understand at least qualitatively and conceptually if not quantitatively.

- The heart, illustrated in the schematic in figure 8.20 above[197] is the "pump" that drives blood through your blood vessels.
- The blood vessels are differentiated into three distinct types:
  - ***Arteries***, which lead strictly ***away from the heart*** and which contain a muscular layer that elastically dilates and contracts the arteries in a synchronous way to help carry the surging waves of blood. This acts as a "shock absorber" and hence reduces the peak systolic blood pressure (see below). As people age, this muscular tissue becomes less elastic – "hardening of the arteries" – as collagen repair mechanisms degrade or plaque (see below) is deposited and systolic blood pressure often increases as a result.

[197] Wikipedia: http://www.wikipedia.org/wiki/Human heart. The diagram itself is borrowed from the wikipedia creative commons, and of course you can learn a lot more of the anatomy and function of the heart and circulation by reading the wikipedia articles on the heart and following links.

Arteries split up the farther one is from the heart, eventually becoming ***arterioles***, the very small arteries that actually split off into capillaries.

- ***Capillaries***, which are a dense network of very fine vessels (often only a single cell thick) that ***deliver oxygenated blood throughout all living tissue*** so that the oxygen can disassociate from the carrying hemoglobin molecules and diffuse into the surrounding cells in systemic circulation, or ***permit the oxygenation of blood*** in pulmonary circulation.
- ***Veins***, which lead strictly ***back to the heart*** from the capillaries. Veins also have a muscle layer that expands or contracts to aid in thermoregulation and regulation of blood pressure as one lies down or stands up. Veins also provide "capacitance" to the circulatory system and store the body's "spare" blood; 60% of the body's total blood supply is usually in the veins at any one time. Most of the veins, especially long vertical veins, are equipped with one-way ***venous valves*** every 4-9 cm that prevent backflow and pooling in the lower body during e.g. diastoli (see below).

  Blood from the capillaries is collected first in ***venules*** (the return-side equivalent of arterioles) and then into veins proper.

- There are two distinct circulatory systems in humans (and in the rest of the mammals and birds):

  - ***Systemic circulation***, where oxygenated blood enters the heart via pulmonary veins *from* the lungs and is pumped at high pressure *into* systemic arteries that deliver it through the capillaries and (deoxygenated) back via systemic veins to the heart.
  - ***Pulmonary circulation***, where deoxgenated blood that has returned *from* the system circulation is pumped *into* pulmonary arteries that deliver it to the lungs, where it is oxygenated and returned to the heart by means of pulmonary veins. These two distinct circulations ***do not mix*** and together, ***form a closed double circulation loop***.

- The heart is the pump that serves ***both systemic and pulmonary circulation***. Blood enters into the ***atria*** and is expelled into the two circulatory system from the ***ventricles***. Systemic circulation enters from the pulmonary veins into the ***left atrium***, is pumped into the ***left ventricle*** through the one-way ***mitral valve***, which then pumps the blood at high pressure into the systemic arteries via the ***aorta*** through the one-way ***aortic valve***. It is eventually returned by the systemic veins (the ***superior and inferior vena cava***) to the ***right atrium***, pumped into the ***right ventricle*** through the one-way ***tricuspid valve***, and then pumped at high pressure into the ***pulmonary artery*** for delivery to the lungs.

  The human heart (as well as the hearts of birds and mammals in general) is thus ***four-chambered*** – two atria and two ventricles. The total resistance of the systemic circulation is generally larger than that of the pulmonary circulation and hence systemic arterial blood pressure must be higher than pulmonary arterial blood pressure in order to maintain the same flow. The left ventricle (primary systemic pump) is thus typically composed of thicker and stronger muscle tissue than the right ventricle. Certain reptiles also have four-chambered hearts, but their pulmonary and systemic circulations are not completely distinct and it is thought that their hearts became four-chambered by a different evolutionary pathway.

- ***Blood pressure*** is generally measured and reported in terms of two numbers:
  - ***Systolic*** blood pressure. This is the ***peak/maximum arterial pressure*** in the wave pulse generated that drives ***systemic circulation***. It is measured in the (brachial artery of the) arm, where it is supposed to be a reasonably accurate reflection of ***peak aortic pressure*** just outside of the heart, where, sadly, it cannot easily be *directly* measured without resorting to invasive methods that are, in fact, used e.g. during surgery.
  - ***Diastolic*** blood pressure. This is the ***trough/minimum arterial pressure*** in the wave pulse of systemic circulation.

  Blood pressure has historically been measured in ***millimeters of mercury*** – torr – in part because until fairly recently a sphygnomometer built using an integrated mercury barometer was by far the most accurate way to measure blood pressure, and it still extremely widely used in situations where high precision is required. Recall that 760 mmHg/torr is *exactly* 1 atm or 101325 Pa.

"Normal" Systolic systemic blood pressure in a very healthy adult can fairly accurately be estimated on the basis of the distance between the heart and the feet; a distance on the order of 1.5 meters leads to a pressure difference of around 0.145 atm or 110 mmHg, and normal systolic is a bit over this to overcome resistance. The "official" normal is currently 120 mmHg/torr[198].

Blood is driven through the relatively high resistance of the capillaries by the ***difference*** in arterial pressure and venous pressure. The venous system is entirely a ***low pressure return***; its peak pressure is typically order of 0.008 bar (6 mmHg/torr). This can be understood and predicted by the mean distance between valves in the venous system – the pressure difference between one valve and another (say) 8 cm higher is approximately $\rho_b g \times 0.08 \approx= 0.008$ bar. However, this pressure is not really static – it varies with the delayed pressure wave (a *wave pulse* – see next chapter) that causes blood to surge its way up, down, or sideways through the veins on its way back to the atria of the heart.

This difference in pressure means one very important thing. If you puncture or sever a vein, blood runs out relatively slowly and is fairly easily staunched, as it is driven by a pressure only a small bit higher than atmospheric pressure. Think of being able to easily plug a small leak in a glass of water (where the fluid height is likely very close to 8-10 cm) by putting your finger over the hole. When blood is drawn, a vein in e.g. your arm is typically tapped, and afterwards the hole almost immediately seals well enough with a simple clot sufficient to hold in the venous pressure.

If you puncture an *artery*, on the other hand, especially a large artery that still has close to the full systolic/diastolic pressure within it, blood ***spurts*** out of it driven by considerable pressure, the pressure one might see at the bottom of a *barrel* or a back yard *swimming pool* a meter and a half or so deep! Not so easy to stop it with light pressure from a finger, or seal up with a clot! It is proportionately more difficult to stop arterial bleeding and one can lose

[198]There is some disagreement here (as of 3/23) – because of side effects of treatment drugs normal is currently a range from 120 to 140, with some physicians electing not to treat until it is over 140.

considerable blood volume in a very short time, leading to a fatal hypotension[199]. Of course if one severs a *large* artery *or* vein (so that clotting has no chance to work) this is a very bad thing, but in general always worse for arteries than for veins, all other things being equal.

An exception of sorts is found in the jugular vein (returning from the brain). It ***has no valves*** because it is, after all, ***downhill*** from the brain back to the heart! As a consequence of this a human who is *inverted* (suspended by their feet, standing on their head) experiences a variety of circulatory problems in their head. Blood pools in the head, neck and brain until blood pressure there (increased by distension of the blood vessels) is enough to lift the blood back up to the heart *without* the help of valves, increasing venous return pressure in the brain itself by a factor of 2 to 4. This higher pressure is transmitted back throughout the brain's vasculature, and, if sustained, can easily cause aneurisms or ruptures of blood vessels (see below)) and death from blood clots or stroke. It can also cause permanent blindness as circulation through the eyes is impaired[200] [201]

Note, however, that blood is pushed through the systemic and pulmonary circulation in *waves* of peak pressure that actually propagate faster than the flow and that the elastic aorta acts like a balloon that is constantly refilled during the left ventricular constraction and "buffers" the arterial pressure – in fact, it behaves a great deal like a ***capacitor in an*** RC ***circuit*** behaves if it is being driven by a series of voltage pulses! Even this model isn't quite adequate. The systolic pulse propagates down through the (elastic) arteries (which dilate and contract to accomodate and maintain it) at a speed ***faster*** than the actual fluid velocity of the blood in the arteries. This effectively rapidly transports a new ejection volume of blood out to the arterioles to replace the blood that made it through the capillaries in the previous heartbeat and maintain a positive pressure difference *across* the capillary system even during diastole so that blood continues to move forward around the circulation loop. This can be mathematically modeled with a very complex circulation model, but is *way* beyond our pay grade at this time – hopefully you are are satisfied with this *heuristic* explanation based on a dynamical "equilibrium" picture of flow plus a bit of intuition as to how an elastic "balloon" can store pressure pulses and smooth out their delivery.

---

[199]In case you don't know – hypo is too small, hyper is too big. You really don't want hypotension (people tend to faint and fall over, a bad thing especially for the elderly) *or* hypertension (which leads to strokes, aneurisms, heart attacks, kidney problems, varicose veins, and more. It even makes your dog want to bite you...

[200]Hanging upside down was actually one of the most brutal forms of medieval execution. Victims retained consciousness throughout because their brain was well-perfused *while* all of these things happened until one of them killed them, or other tortures were applied that did the job. Humans historically have used knowledge for evil as often as they ever use it for good.

[201]"You are old, Father William," the young man said,
"And your hair has become very white;
And yet you incessantly stand on your
Do you think, at your age, it is right?"

"In my youth," Father William replied to his son,
"I feared it might injure the brain;
But now that I'm perfectly sure I have none,
Why, I do it again and again."

Alice in Wonderland, by Lewis Carroll

Not really that good an idea...

There is a systolic peak in the blood pressure delivered to the pulmonary system as well, but it is ***much lower*** than the systolic systemic pressure, because (after all) the lungs are at pretty much the same height as the heart and the vessels leading to and from it are also quite short (resistance is proportional to *length*, among other things, recall). The pressure difference needs only to be high enough to maintain the same average flow as that in the systemic circulation through the much lower resistance of the capillary network in the lungs. Typical healthy normal resting pulmonary systole pressures are around 15 mmHg; pressures higher than 25 mmHg are likely to be associated with ***pulmonary edema***, the pooling of fluid in the interstitial spaces of the lungs. This is a bad thing and is even fluidic, but is beyond the scope of this topic in the chapter.

The return pressure in the pulmonary veins (to the right atrium, recall) is 2-15 mmHg, roughly consistent with central (systemic) venous pressure at the left atrium but with a larger variation.

The actual rhythm of the heartbeat is as follows. Bearing in mind that it is a continuous cycle, "first" blood from the last beat, which has accumulated in the left and right atrium during the heart's resting phase between beats, is expelled through the mitral and tricuspid valves from the left and right atrium respectively. The explusion is accomplished by the contraction of the atrial wall muscles, and the backflow of blood into the venous system is (mostly) prevented by the valves of the veins although there can be some small regurgitation. This expulsion (which is nearly simultaneous, with mitral barely leading tricuspid) pre-compresses the blood in the ventricles and is the first "lub" one hears in the "lub-*dub*" of the heartbeat observed with a stethoscope.

Immediately following this a (nearly) simultanous contraction of the ventricular wall muscles causes the expulsion of the blood from the left and right ventricles through the semilunar valves into the aorta and pulmonary arteries, respectively. As noted above, the left-ventricular wall is thicker and stronger and produces a substantially higher peak systolic pressure in the aorta and systemic circulation than the right ventricle produces in the pulmonary arteries. The elastic aorta distends (reducing the peak systemic pressure as it stores the energy produced by the heartbeak) and partially sustains the higher pressure across the capillaries throughout the resting phase, driving the aforementioned wave pulse in the arterial system. This is the "dub" in the "lub-*dub*" of the heartbeat.

This isn't anywhere near *all* of the important physics in the circulatory system, but it should be enough for you to be able to understand the basic plumbing well enough to learn more

later[202]. Let's look at a few very important examples of how things go *wrong* in the circulatory system and see how the physics of fluids helps you understand and detect them!

Example 8.5.1: Atherosclerotic Plaque Partially Occludes a Blood Vessel

Humans are not yet evolved to live 70 or more years. Mean life expectancy as little as a hundred years ago was in the mid-50's, *if* you only average the people that survived to age 15

[202]...And understand why we spent time learning to solve first order differential equations this semester so that we can understand RC circuits in detail next semester so we can think *back* and further understand the remarks about how the aorta acts like a capacitor this semester.

– otherwise it was in the 30's! The average age when a woman bore her first child throughout most of the period we have been considered "human" has been perhaps 14 or 15, and a woman in her thirties was often a grandmother. Because evolution works best if parents don't hang around *too* long to compete with their own offspring, we are very likely *evolved to die* somewhere around the age of 50 or 60, three to four generations (old style) after our own birth. Humans begin to really experience the effects of aging – failing vision, incipient cardiovascular disease, metabolic slowing, greying hair, wearing out teeth, cancer, diminished collagen production, arthritis, around age 45 (give or take a few years), and it once it starts it just gets worse. Old age *physically* sucks, I can say authoritatively as I type this peering through reading glasses with my mildly arthritic fingers over my gradually expanding belly at age ~~56~~,~~59~~,~~61~~...*68* (tomorrow!).

One of the many ways it sucks is that the 40's and 50's is where people usually show the first signs of *cardiovascular disease*, in particular ***atherosclerosis***[203] – granular deposits of fatty material called ***plaques*** that attach to the walls of e.g. arteries and gradually thicken over time, generally associated with high blood cholesterol and lipidemia. The risk factors for atherosclerosis form a list as long as your arm and its fundamental causes are not well understood, although they are currently believed to form as an inflammatory response *involving* surplus low density lipoproteins (one kind of cholesterol) in the blood[204].

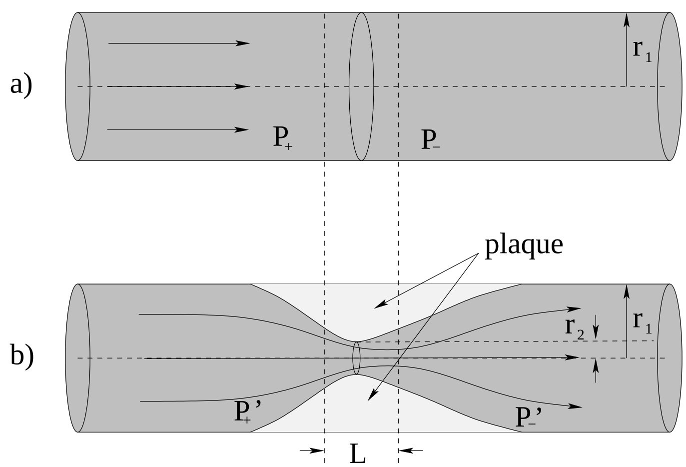


Figure 8.21: Two "identical" blood vessels with circular cross-sections, one that is clean (of radius r_1) and one that is perhaps 90% occluded by plaque that leaves an aperture of radius $r_2 < r_1$ in a region of some length L.

Our purpose, however, is not to think about causes and cures but instead what *fluid physics* has to say about the disorder, its diagnosis and effects. In figure 8.21 two arteries are illustrated. Artery a) is "clean", has a radius of r_1, and (from the Poiseuille Equation above) has a very *low resistance* to any given flow of blood. Because R_a over the length L is low, there is very little pressure drop between P_+ and P_- on the two sides of any given stretch of length L. The velocity profile of the fluid is also more or less uniform in the artery, slowing a bit near the

[203] Wikipedia: http://www.wikipedia.org/wiki/Atherosclerosis. As always, there is far, far more to say about this subject than I can cover here, all of it interesting and capable of helping you to select a lifestyle that prolongs a high quality of life.

[204] Although taking medications that dramatically lower LDL levels *alone* have almost no effect on morbidity or mortality unless they are *statins*, go figure.

walls but generally moving smoothly throughout the entire cross-section.

Artery b) has a significant deposit of atherosclerotic plaques that have coated the walls and reduced the effective radius of the vessel to $\sim r_2$ over an extended length L. The vessel is perhaps 90% occluded – only 10% of its normal cross-sectional area is available to carry fluid.

We can now easily understand several things about this situation. First, if the total *flow* in artery b) is still being maintained at close to the levels of the flow in artery a) (so that tissue being oxygenated by blood delivered by this artery is not being critically starved for oxygen yet) the ***fluid velocity in the narrowed region is ten times higher than normal!*** Since the Reynolds number for blood flowing in primary arteries is normally around 1000 to 2000, increasing v by a factor of 10 increases the Reynolds number by a factor of 10, causing the flow to become *turbulent* in the obstruction. This tendency is even more pronounced than this figure suggests – I've drawn a nice symmetric occlusion, but the atheroma (lesion) is more likely to grow predominantly on one side and irregular lesions are more likely to disturb laminar flow even for smaller Reynolds numbers.

This turbulence provides the basis for one method of possible detection and diagnosis – you can *hear* the turbulence (with luck) through the stethoscope during a physical exam. Physicians get a lot of practice listening for turbulence since turbulence produced by *artificially* restricting blood flow in the brachial artery by means of a constricting cuff is basically what one listens for when taking a patient's blood pressure. It really shouldn't be there, especially during diastole, the rest of the time.

Next, consider what the vessel's resistance across the lesion of length L should do. Recall that $R \propto 1/A^2$. That means that the resistance is at least ***100 times larger*** than the resistance of the healthy artery over the same distance. In truth, it is almost certainly much greater than this, because as noted, one has converted to turbulent flow and our expression for the resistance assumed *laminar* flow. It's the metaphorical difference between people crowded into a large room moving through a door in an orderly line (and speeding up once they are through to move down the hall, perhaps during a fire drill) and the same crowd of people panicked and bouncing off of the walls to constantly disturb forward progress, even in the halls. Turbulence produces *quadratic* drag resistance.

Frankly, work in biophysics remains to be done here – nearly everything I found on the subject continues to assume laminar flow in estimates of increased resistance, increased pressure difference, decreased flow, but surely this is a lower bound likely off by some *further* increase in the power of A in the denominator!

Regardless, a hundredfold to thousandfold increase in the resistance of the segment means that *either* the fluid flow itself will be reduced, assuming a constant upstream pressure, *or* the pressure upstream will increase – perhaps substantially – to maintain adequate flow and perfusion. In practice a certain amount of both can occur – the stiffening of the artery due to the lesion and an increased resting heart rate[205] can raise systolic blood pressure, which tends to maintain flow, but as narrowing proceeds it cannot raise it *enough* to compensate, especially not without causing far greater damage somewhere else.

At some point, the tissue downstream from the occluded artery begins to suffer from lack of oxygen, especially during times of metabolic stress. If the tissue in question is in a leg or

[205] Among many other things. High blood pressure is extremely multifactorial.

an arm, weakness and pain may result, not good, but arguably recoverable. If the tissue in questions is ***the heart itself*** or ***the lungs*** or ***the brain*** (or, really, any of your mission-critical organs) this is ***very bad indeed***. The failure to deliver sufficient oxygen to the heart over the time required to cause actual death of heart muscle tissue is (one of the things that is) commonly known as a heart attack. The same failure in major blood vessels that supply the brain is called a stroke. The heart and brain have very limited ability to regrow damaged tissue after either of these events. Occlusion and hardening of the pulmonary arteries can lead to pulmonary hypertension, which in turn (as already noted) can lead to pulmonary edema and a variety of associated problems.

Example 8.5.2: Aneurisms

An aneurism is basically the *opposite* of an atherosclerotic lesion. A portion of the walls of an artery or, less commonly, a vein *thins* and begins to ***dilate*** or stretch in response to the pulsing of the systolic wave. Once the artery has "permanently" stretched along some short length to a larger radius than the normal artery on either side, a nasty feedback mechanism ensues. Since the cross-sectional area of the dilated area is ***larger***, fluid flow there ***slows*** from conservation of flow. At the same time, the pressure in the dilated region must *increase* according to Bernoulli's equation – the pressure increase is responsible for slowing the fluid as it enters the aneurism and re-accelerating it back to the normal flow rate on the far side.

The higher pressure, of course, then makes the already weakened arterial wall stretch more, which dilates the aneurism more, which slows the blood more which increases the pressure, until some sort of limit is reached: extra pressure from surrounding tissue serves to reinforce the artery and keeps it from continuing to grow or the aneurism *ruptures*, spilling blood into surrounding (low pressure) tissue with every heartbeat. While there aren't a lot of places a ruptured aneurism is *good*, in the brain it is very bad magic, causing the same sort of damage as a stroke as the increased pressure in the tissue surrounding the rupture becomes so high that normal capillary flow through the tissue is compromised. Aortic aneurisms are also far from unknown and, because of the high blood flow directly from the heart, the rupture of an aortic aneurism can cause death in as little as *a few minutes* as one bleeds, under substantial pressure, internally.

Example 8.5.3: The Giraffe

"The Giraffe" isn't really an example *problem*, it is more like a nifty/cool True Fact but I haven't bothered to make up a nifty cool true fact header for the book (at least not yet). Full grown adult giraffes are animals (you may recall, or not[206]) that stand roughly 5 meters high.

Because of their height, giraffes have a uniquely evolved circulatory system[207] . In order to drive blood from its feet up to its brain, especially in times of stress when it is e.g. running, it's heart has to be able to maintain a pressure difference of close to *half an atmosphere of pressure* (using the rule that 10 meters of water column equals one atmosphere of pressure

[206]Wikipedia: http://www.wikipedia.org/wiki/Giraffe. And Wikipedia stands ready to educate you further, if you have never seen an actual Giraffe in a zoo and want to know a bit about them

[207]Wikipedia: http://www.wikipedia.org/wiki/Giraffe#Circulatory_system.

difference and assuming that blood and water have roughly the same density). A giraffe heart is correspondingly huge: roughly 60 cm long and has a mass of around 10 kg in order to accomplish this.

When a giraffe is erect, its cerebral blood pressure is “normal” (for a giraffe), but when it bends to drink, its head goes down to the ground. This rapidly increases the blood pressure being delivered by its heart to the brain by 50 kPa or so. It has evolved a complicated set of pressure controls in its neck to reduce this pressure so that it doesn’t have a brain aneurism every time it gets thirsty!

Giraffes, like humans and most other large animals, have a second problem. The heart doesn’t maintain a *steady* pressure differential in and of itself; it expels blood in *beats*. In between contractions that momentarily increase the pressure in the arterial (delivery) system to a ***systolic peak*** that drives blood over into the venous (return) system through capillaries that either oxygenate the blood in the pulmonary system or give up the oxygen to living tissue in the rest of the body, the arterial pressure decreases to a ***diastolic minimum***.

Even in relatively short (compared to a giraffe!) adult humans, the blood pressure differential between our nose and our toes is around 0.16 bar, which not-too-coincidentally (as noted above) is equivalent to the 120 torr (mmHg) that constitutes a fairly “normal” systolic blood pressure. The normal diastolic pressure of 70 torr (0.09 bar) is insufficient to keep blood in the venous system from “falling back” out of the brain and pooling in and distending the large veins of the lower limbs.

To help prevent that, long (especially vertical) veins have ***one-way valves*** that are spaced roughly every 4 to 8 cm along the vein. During systoli, the valves open and blood pulses forward. During diastoli, however, the valves *close* and *distribute the weight of the blood in the return system* to $\sim$6-8 cm segments of the veins while preventing backflow. The pressure differential across a valve and supported by the smooth muscle that gives tone to the vein walls is then just the pressure accumulated across 6 cm (around 5 torr).

Humans get ***varicose veins***[208] when these valves *fail* (because of gradual loss of tone in the veins with age, which causes the vein to swell to where the valve flaps don’t properly meet, or other factors). When a valve fails, the next-lower valve has to support twice the pressure difference (say 10 torr) which in turn swells that vein close to the valve (which can cause it to fail as well) passing three times this differential pressure down to the next valve and so on. Note well that there are two aspects of this extra pressure to consider – one is the increase in pressure differential across the valve, but perhaps the greater one is the increase in pressure differential between the inside of the vein and the outside tissue. The latter causes the vein to *dilate* (swell, increase its radius) as the tissue stretches until its tension can supply the pressure needed to confine the blood column. As always, swelling causes hemodynamic pressure increases where the blood slows, making things incrementally worse.

Opposing this positively fed-back tendency to dilate, which compromises valves, which in turn increases the dilation to eventual destruction are things like muscle use (contracting surrounding muscles exerts extra pressure on the outside of veins and hence decreases the pressure differential and stress on the venous tissue), general muscle and skin tone (the skin and surrounding tissue helps maintain a pressure outside of the veins that is already higher

[208] Wikipedia: http://www.wikipedia.org/wiki/Varicose Veins.

than ambient air pressure, and keeping one's blood pressure under control, as the diastolic pressure sets the scale for the venous pressure during diastoli and if it is high then the minimum pressure differential across the vein walls will be correspondingly high. Elevating one's feet can also help, exercise helps, wearing support stockings that act like a second skin and increase the pressure outside of the veins can help.

Carrying the extra pressure below compromised valves nearly all of the time, the veins – especially those near the skin and hence minimally supported by surrounding tissue – gradually dilate until they are many times their normal diameter, and significant amounts of blood pool in them – these "ropy", twisted, fat, deformed veins that not infrequently visibly pop up out of the skin in which they are embedded are the varicose veins.

Naturally, giraffes have this problem in spades. The pressure in their lower extremities, even allowing for their system of valves, is great enough to rupture their capillaries. To keep this from happening, the skin on the legs of a giraffe is among the thickest and strongest found in nature – it functions like an astronaut's "g-suit" or a permanent pair of support stockings, maintaining a high baseline pressure in the *tissue* of the legs outside of the veins and capillaries, and thereby reducing the pressure differential.

Another cool fact about giraffes – as noted above, they pretty much live with "high blood pressure" – their normal pressure of 280/180 torr (mmHg) is 2-3 times that of humans (because their height is 2-3 times that of humans) in order to keep their brain perfused with blood. This pressure has to further elevate when they e.g. run away from predators or are stressed. Older adult giraffes have a tendency to ***die of a heart attack*** if they run for too long a time, so zoos have to take care to avoid stressing them if they wish to capture them!

Finally, giraffes splay their legs when they drink so that they reduced the pressure differential the peristalsis of their gullet has to maintain to pump water up and over the hump down into their stomach. Even this doesn't completely exhaust the interesting list of giraffe facts associated with their fluid systems. Future physicians would be well advised to take a closer peek at these very large mammals (as well as at elephants, who have many of the same problems but very different evolutionary adaptations to accommodate them) in order to gain insight into the complex fluid dynamics to be found in the human body!

Homework for Week 8

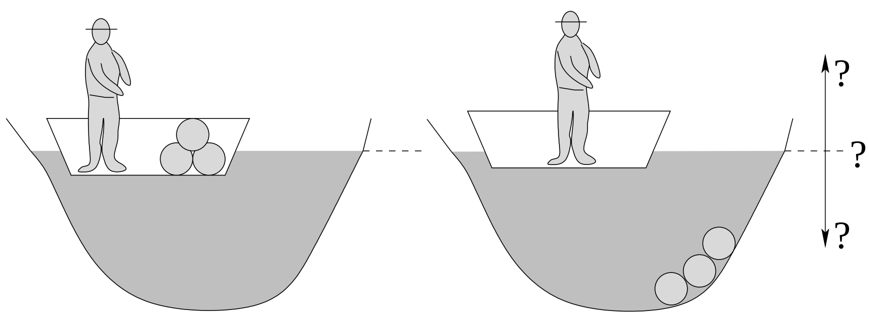
Problem 1.

Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

Problem 2.

A small boy is riding in a minivan with the windows closed, holding a helium balloon. The van goes around a corner to the left. Does the balloon swing to the left, still pull straight up, or swing to the right as the van swings around the corner?

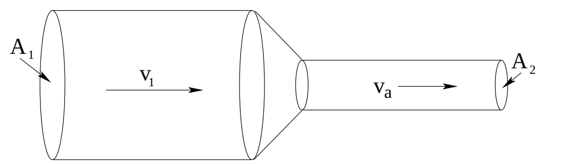
Problem 3.



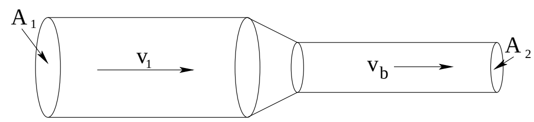
A person stands in a boat floating on a pond and containing several large, round, rocks. He throws the rocks out of the boat so that they sink to the bottom of the pond. The water level of the pond will:

- ☐ Remain unchanged.
- ☐ Rise a bit.
- ☐ Fall a bit.
- ☐ Can't tell from the information given (it depends on, for example, the shape of the boat, the mass of the person, whether the pond is located on the Earth or on Mars...).

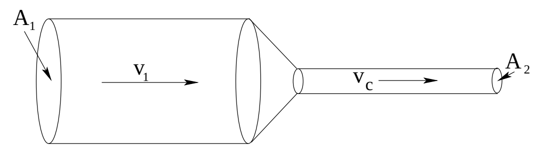
Problem 4.



a) $A_1 = 30\ \text{cm}^2$, $v_1 = 3$ cm/sec, $A_2 = 6\ \text{cm}^2$



b) $A_1 = 10\ \text{cm}^2$, $v_1 = 8$ cm/sec, $A_2 = 5\ \text{cm}^2$

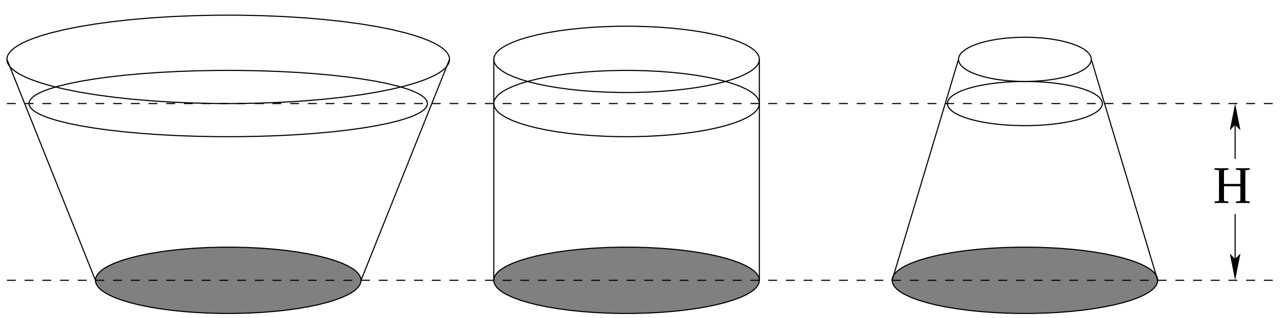


c) $A_1 = 20\ \text{cm}^2$, $v_1 = 3$ cm/sec, $A_2 = 3\ \text{cm}^2$

In the figure above three different pipes are shown, with cross-sectional areas and flow speeds as shown. **Rank the three diagrams** a, b, and c in the **order of the speed of the outgoing flow** (fill the appropriate letters in the provided boxes):

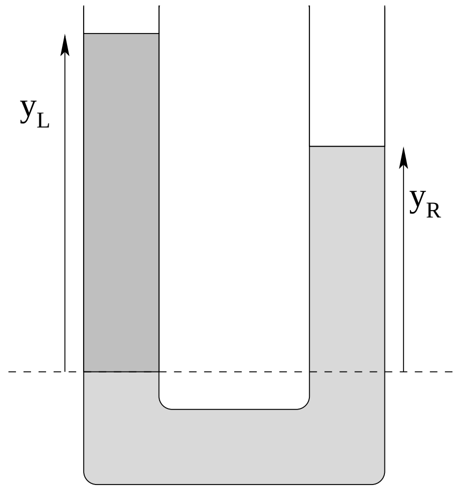
$$v_{\square} > v_{\square} > v_{\square}$$

Problem 5.



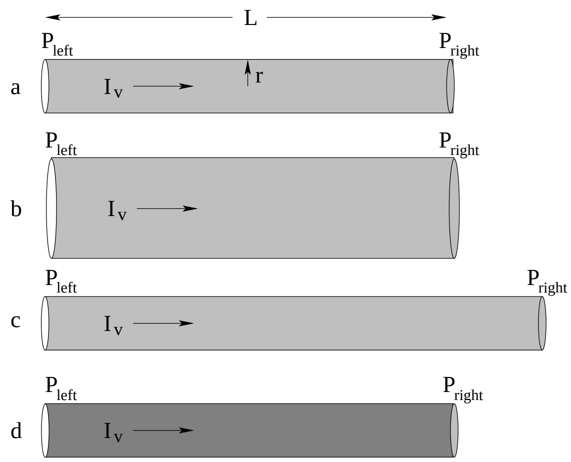
In the figure above three flasks are drawn that have the same (shaded) cross sectional area of the bottom. The depth of the water in all three flasks is H, and so the pressure at the bottom in all three cases is the same. Explain how the force exerted **by the fluid** on the circular bottom can be the same for all three flasks when all three flasks contain different weights of water.

Problem 6.



A vertical U-tube open to the air at the top is filled with oil (density ρ_o) on one side and water (density ρ_w) on the other, where $\rho_o < \rho_w$. Find y_L, the height of the column on the left, in terms of the densities, g, and y_R as needed. Clearly label the oil and the water on the diagram above and show all reasoning including the basic principle(s) upon which your answer is based.

Problem 7.



In this class we usually idealize fluid flow by neglecting resistance (drag) and the viscosity of the fluid as it passes through cylindrical pipes so we can use the Bernoulli equation. As discussed in class, however, it is often necessary to have at sound conceptual understanding of at least the *qualitative effects* and *scaling* of the effects of resistance and viscosity.

Use ***Poiseuille's Law*** and the concepts of ***resistance, pressure, and flow*** to ***qualitatively and conceptually*** determine the change in blood pressure differential across a segment of blood vessel assume that the flow (required to maintain tissue perfusion) remains constant[209]. In all the figures

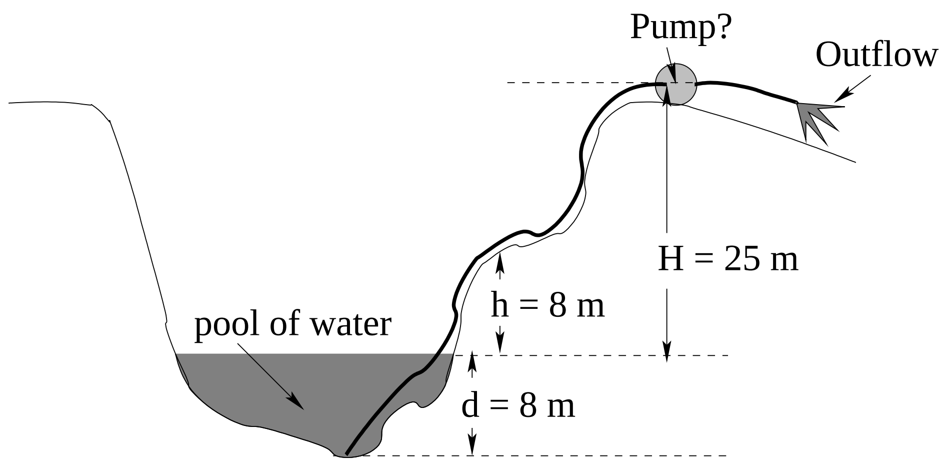
$$\Delta P_i = P_{\text{left}} - P_{\text{right}} \quad \text{for figure} \quad i = a, b, c, d$$

and you should fill in the provided box (or a copy on your own paper) with $<, >, =$. Figure a) is the "standard" for answering b), c), and d).

a)	Blood flows at a rate I_v horizontally through a blood vessel with constant radius r and some length L.	ΔP_a	$\square$	0
b)	I_v and L remain the same, but r increases (relative to a)).	ΔP_b	$\square$	P_a
c)	I_v and r remain the same, but L increases (relative to a)).	ΔP_c	$\square$	P_a
d)	I_v, r and L remain the same, but μ (dynamical viscosity) increases (relative to a)).	ΔP_d	$\square$	P_a

[209]https://www.cvphysiology.com/Hemodynamics/H011 Blood viscosity is *complicated* – it is a non-Newtonian fluid – but nevertheless is a key component of understanding the circulatory system and cardiovascular disease! Viscosity and the radius r of blood vessels can be regulated or altered (within limits) by drugs, disease, diet and exercise – all of which have a ***completely understandable*** effect upon blood pressure and perfusion of cells based on the ideas reviewed in this problem.

Problem 8.

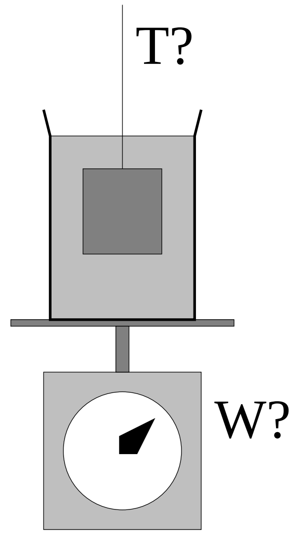


A pump is a machine that can maintain a pressure differential between its two sides. A particular pump that can maintain a pressure differential of as much as 10 atmospheres of pressure between the low pressure side and the high pressure side is being used on a construction site.

a) Your construction boss has just called you into her office to help her figure out why they aren't getting any water out of the pump on top of the $H = 25$ meter high cliff shown above. Examine the schematic above and show why it cannot possibly deliver water to the pump at the top of the cliff. Your explanation should include an invocation of the appropriate physical law(s) and an explicit calculation of the highest distance the a pump *could* lift water in this arrangement. Why is the notion that the pump "sucks water up" misleading? What really moves the water up?

b) If you answered a), you get to keep your job. If you answer b), you might even get a raise (or at least, get full credit on this problem)! Tell your boss where this single pump should be located to move water up to the top and show (draw a picture of) how it should be hooked up. There are at least a couple of straightforward ways to answer this question.

Problem 9.



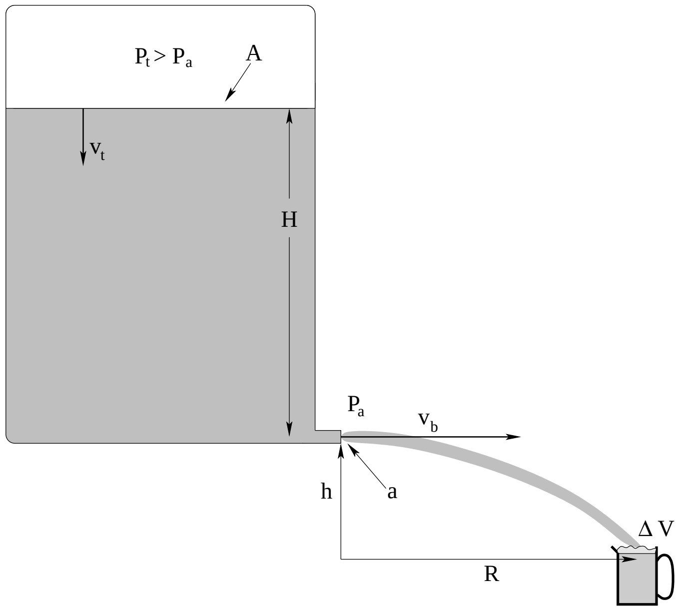
A block of density ρ and volume V is suspended by a thin thread and is immersed completely in a jar of oil (density $\rho_o < \rho$) that is resting on a scale as shown. The total mass of the oil and jar (alone) is M.

a) What is the buoyant force exerted by the oil on the block?

b) What is the tension T in the thread?

c) What weight W does the scale read?

Problem 10.



In the figure above, a CO_2 cartridge is used to maintain a pressure P on top of the beer in a beer keg, which is full up to a height H above the tap at the bottom (which is obviously open to normal air pressure) a height h above the ground. The keg has a cross-sectional area A at the top. Somebody has pulled the tube and valve off of the tap (which has a cross sectional area of a) at the bottom.

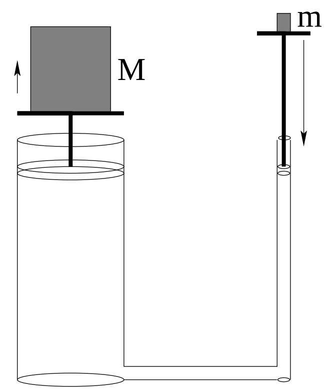
a) Find the speed v_b with which the beer emerges from the tap. You may use the approximation $A \gg a$, but please do so only at the end. Assume laminar flow and no resistance.

b) Find the value of R at which you should place a pitcher (initially) to catch the beer.

c) Assuming the volume of the pitcher ΔV is much less than the volume of the keg and that the beer does not foam up, how long will it take to fill the pitcher?

d) Evaluate the answers to a)–c) for $A = 0.25$ m^2, $P = 2$ atmospheres, $a = 0.25$ cm^2, $H = 50$ cm, $h = 1$ meter and $\rho_{\text{beer}} = 1000$ kg/m^3 (the same as water). The volume of the pitcher is reasonably $\Delta V = 2000$ cm^3.

Problem 11.

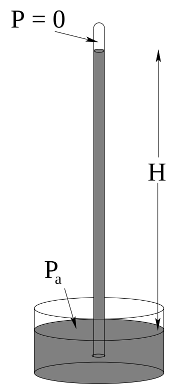


In a hydraulic lift apparatus, a pair of coupled cylinders are filled with an *incompressible*, very light fluid (assume that the mass of the fluid is 'zero' compared to everything else).

a) If the mass M on the left is 1000 kilograms, the cross-sectional area of the left piston is 100 cm^2, and the cross sectional area of the right piston is 1 cm^2, what mass m should one place on the right for the two objects to be in balance?

b) Suppose one pushes the right piston down a distance of one meter. How much does the mass M rise?

Problem 12.



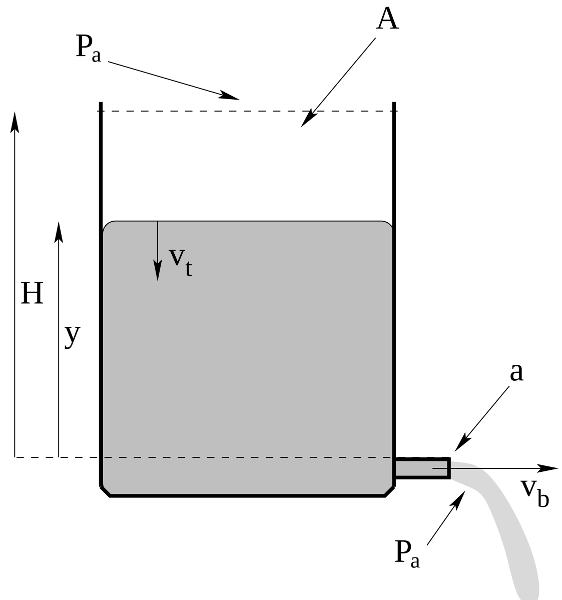
The idea of a barometer is a simple one. A tube filled with a suitable liquid is inverted into a reservoir. The tube empties (maintaining a seal so air bubbles cannot get into the tube) until the static pressure in the liquid is in balance with the *vacuum* that forms at the top of the tube and the ambient pressure of the surrounding air on the fluid surface of the reservoir at the bottom.

a) Suppose the fluid is water, with $\rho_w = 1000$ kg/m^3. Approximately how high will the water column be? Note that water is not an ideal fluid to make a barometer with because of the height of the column necessary and because of its annoying tendency to boil at room temperature into a vacuum.

b) Suppose the fluid is mercury, with a specific gravity of 13.6. How high will the mercury column be? Mercury, as you can see, *is* nearly ideal for fluids-pr-compare-barometers except for the minor problem with its extreme toxicity and high vapor pressure.

Fortunately, there are many other ways of making good barometers that don't involve heavy, toxic fluids or light, volatile fluids!

Advanced Problem 13.



In the figure to the left, a large drum of water (density ρ_w) is open at the top and filled up to an initial height H above a tap at the bottom (which is also open to normal air pressure P_a of 1 atmosphere). The drum has a cross-sectional area A at the top and the tap has a cross sectional area of a, where you may assume $a \ll A$. You may also assume laminar flow and neglect resistance and/or drag.

a) ***Make a guess*** as to how long it will take water to flow out of the tank by ***using dimensional analysis*** and Torrecelli's Law.

b) *Find* the speed v_b with which the water emerges from the b(ottom) tap when the surface of the water is at an arbitrary height $y < H$, *without* directly using Torrecelli's Law. Use this to determine the speed of descent of the surface at the t(op) $v_t = -\frac{dy}{dt}$ and write the result as a (very simple, integrable) differential equation.

c) Solve (integrate) the differential equation you obtained in a) and use algebra to find the time required to empty the tank t_e to a height just above the top edge of the bottom tap, starting with it completely full ($y(0) = H$, $y(t_e) = 0$). Compare your answer to b) to your answer and to the time it takes a mass to fall a height y in a uniform gravitational field. Does the your answer also make dimensional and physical sense?.

d) Evaluate the answer to c) for $A = 0.125$ m^2, $a = 0.5$ cm^2 and $y(0) = H = 50$ cm.

NOTE WELL: Start Review for Final!

At this point we are roughly four "weeks" out from our final exam[210]. I thus **strongly suggest** that you devote any extra time you have to a gradual slow review of all of the basic physics from the first half of the course. Make sure that you still remember and understand all of the basic principles of Newton's Laws, work and energy, momentum, rotation, torque and angular momentum. Look over your old homework and quiz and hour exam problems, review problems out of your notes, and look for help with any ideas that still aren't clear and easy!

[210]...which might be only *one and a half* weeks out in a summer session!

Week 9: Oscillations

1.16: Oscillation Summary

- Springs obey Hooke's Law: $\vec{F} = -k\vec{x}$ (where k is called the *spring constant*. A perfect spring (with no damping or drag force) produces perfect harmonic oscillation, so this will be our archetype.
- A pendulum (as we shall see) has a restoring force or torque proportional to displacement for *small* displacements but is much too complicated to treat in this course for large displacements. It is a simple example of a problem that oscillates harmonically for small displacements but *not* harmonically for large ones.
- An oscillator can be **damped** by dissipative forces such as friction and viscous drag. A damped oscillator can have exhibit a variety of behaviors depending on the relative strength and form of the damping force, but for one special form it can be easily described.
- An oscillator can be ***driven*** by e.g. an external harmonic driving force that may or may not be at the same frequency (in ***resonance*** with the natural frequency of the oscillator.
- The equation of motion for any (undamped) harmonic oscillator is the same, although it may have different dynamical variables. For example, for a spring it is:
$$\frac{d^2x}{dt^2} + \frac{k}{m}x = \frac{d^2x}{dt^2} + \omega^2 x = 0 \tag{9.1}$$
where for a simple pendulum (for small oscillations) it is:
$$\frac{d^2\theta}{dt^2} + \frac{g}{\ell}x = \frac{d^2\theta}{dt^2} + \omega^2\theta = 0 \tag{9.2}$$
(In this latter case ω is the ***angular frequency*** of the oscillator, not the angular velocity of the mass $d\theta/dt$.)
- The general solution to the equation of motion is:
$$x(t) = A\cos(\omega t + \phi) \tag{9.3}$$
where $\omega = \sqrt{k/m}$ and the amplitude A (units: length) and phase ϕ (units: dimensionless/radians) are the constants of integration (set from e.g. the initial conditions). Note that we alter the variable to fit the specific problem – for a pendulum it would be:
$$\theta(t) = \Theta\cos(\omega t + \phi) \tag{9.4}$$
with $\omega = \sqrt{g/\ell}$, where the angular amplitude Θ now has units of radians.

- The velocity of the mass attached to an oscillator is found from:

$$v(t) = \frac{dx}{dt} = -A\omega \sin(\omega t + \phi) = -V \sin(\omega t + \phi) \tag{9.5}$$

(with $V = v_{\rm max} = A\omega$).

- From the velocity equation above, we can easily find the kinetic energy as a function of time:

$$K(t) = \frac{1}{2}mv^2 = \frac{1}{2}mA^2\omega^2 \sin^2(\omega t + \phi) = \frac{1}{2}kA^2 \sin^2(\omega t + \phi) \tag{9.6}$$

- The potential energy of an oscillator is found by integrating:

$$U(x) = -\int_0^x -kx' dx' = \frac{1}{2}kx^2 = \frac{1}{2}kA^2 \cos^2(\omega t + \phi) \tag{9.7}$$

if we use the (usual but not necessary) convention that $U(0) = 0$ when the mass is at the equilibrium displacement, $x = 0$.

- The total mechanical energy is therefore obviously a constant:

$$E(t) = \frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{1}{2}kA^2 \sin^2(\omega t + \phi) + \frac{1}{2}kA^2 \cos^2(\omega t + \phi) = \frac{1}{2}kA^2 \tag{9.8}$$

- As usual, the relation between the angular frequency, the regular frequency, and the period of the oscillator is given by:

$$\omega = \frac{2\pi}{T} = 2\pi f \tag{9.9}$$

(where $f = 1/T$). SI units of frequency are ***Hertz*** – cycles per second. Angular frequency has units of radians per second. Since both cycles and radians are dimensionless, the units themselves are dimensionally inverse seconds but they are (obviously) related by 2π radians per cycle.

- A (non-ideal) harmonic oscillator in nature is almost always ***damped*** by friction and drag forces. If we assume damping by viscous drag in (low Reynolds number) laminar flow – not unreasonable for smooth objects moving in a damping fluid, if somewhat itself idealized – the equation of motion becomes:

$$\frac{d^2x}{dt^2} + +\frac{b}{m}\frac{dx}{dt} + \frac{k}{m}x = 0 \tag{9.10}$$

- The solution to this equation of motion is:

$$x_\pm(t) = A_\pm e^{\frac{-b}{2m}t} \cos(\omega' t + \phi) \tag{9.11}$$

where

$$\omega' = \omega_0 \sqrt{1 - \frac{b^2}{4km}} \tag{9.12}$$

9.1: The Simple Harmonic Oscillator

Oscillations occur whenever a force exists that pushes an object back towards a stable equilibrium position whenever it is displaced from it. Such forces abound in nature – things are held together in structured form *because* they are in stable equilibrium positions and when they are disturbed in certain ways, they oscillate.

When the displacement from equilibrium is *small*, the restoring force is often *linearly* related to the displacement, at least to a good approximation. In that case the oscillations take on a special character – they are called *harmonic* oscillations as they are described by harmonic functions (sines and cosines) known from trigonometery.

In this course we will study **simple harmonic oscillators**, both with and without damping (and harmonic driving) forces. The principle examples we will study will be masses on springs and various penduli.

9.1.1: The Archetypical Simple Harmonic Oscillator: A Mass on a Spring

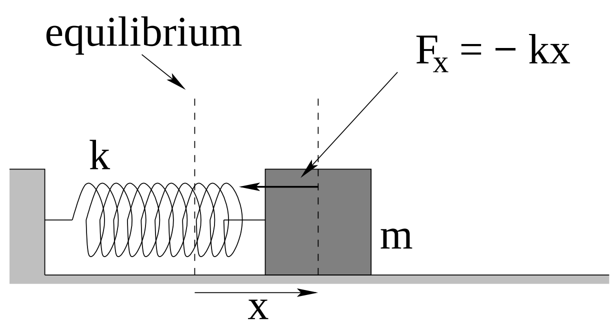


Figure 9.1: A mass on a spring is displaced by a distance x from its equilibrium position. The spring exerts a force $F_x = -kx$ on the mass (Hooke's Law).

Consider a mass m attached to a perfect spring (which in turn is affixed to a wall as shown in figure 9.1. The mass rests on a surface so that gravitational forces are cancelled by the normal force and are hence irrelevant. The mass will be displaced only in one direction (call it x) and otherwise constrained so that motion in or out of the plane is impossible and no drag or frictionless forces are (yet) considered to be relevant.

We know that the force exerted by a perfect spring on a mass stretched a distance x from its equilibrium position is given by ***Hooke's Law:***

$$F_x = -kx \tag{9.13}$$

where k is the ***spring constant*** of the spring. This is a ***linear restoring force***, and (as we shall see) is highly characteristic of the restoring forces exerted by *any* system around a point of stable equilibrium.

Although thus far we have avoided trying to determine the most general motion of the mass, it is time for us to tackle that chore. As we shall see, the motion of an undamped simple harmonic oscillator is *very easy to understand* in the ideal case, and easy enough to understand qualitatively or semi-quantitatively that it serves as an excellent springboard to understanding many of the properties of bulk materials, such as ***compressibility, stress, and strain***.

We begin, as one might expect, with Newton's Second Law, used to obtain the (second order, linear, homogeneous, differential) equation of motion for the system. Note well that although this sounds all complicated and everything – like "real math" – we've been solving second order differential equations from day one, so they *shouldn't* be intimidating. Solving the equation of motion for the simple harmonic oscillator isn't *quite* as simple as just integrating twice, but as we will see neither is it all that difficult.

Hooke's Law combined with Newton's Second Law can thus be written and massaged algebraically as follows:

$$\begin{aligned} ma_x = m\frac{d^2x}{dt^2} &= F_x = -kx \\ m\frac{d^2x}{dt^2} + kx &= 0 \\ \frac{d^2x}{dt^2} + \frac{k}{m}x &= 0 \\ \frac{d^2x}{dt^2} + \omega^2 x &= 0 \end{aligned} \tag{9.14}$$

where we have defined the ***angular frequency*** of the oscillator,

$$\omega^2 = k/m \tag{9.15}$$

This must have units of ***inverse time squared*** (why?). We will momentarily justify this identification, but it won't hurt to start learning it now.

Equation 9.14 (with the ω^2) is the ***standard harmonic oscillator differential equation of motion*** (SHO ODE). As we'll soon see with quite a few examples and an algebraic argument, we can put the equation of motion for ***many*** systems into this form for at least small displacements from a stable equilibrium point. If we can properly solve it *once and for all* now, whenever we can put an equation of motion into this form in the future we can just ***read off the solution*** by ***identifying similar quantities*** in the equation.

To solve it[211], we note that it basically tells us that $x(t)$ must be a function that has a ***second derivative proportional to the function itself***.

We know at least three functions whose second derivatives are proportional to the functions themselves: cosine, sine and exponential. In this particular case, we *could* guess cosine or sine and we would get a perfectly reasonable solution. The bad thing about doing this is that the solution methodology would not generalize at all – it wouldn't work for first order, third order, or even *general* second order ODEs. It would give us a solution to the SHO problem (for example) but *not* allow us to solve the *damped* SHO problem or *damped, driven* SHO problems we investigate later this week. For this reason, although it is a bit more work now, we'll search for a solution ***assuming that it has an exponential form***.

---

[211] Not only it, but *any* homogeneous linear Nth order ordinary differential equation – the method can be applied to first, third, fourth, fifth... order linear ODEs as well.

Note Well!

If you are *completely panicked* by the following solution, if thinking about trying to understand it makes you feel sick to your stomach, you can probably *skip ahead to the next section* (or rather, begin reading again at the end of this chapter after the ***real*** solultion is obtained).

There is a price you will pay if you do. You will never understand where the solution comes from or how to solve the slightly more difficult damped SHO problem, and will therefore have to ***memorize the solutions***, unable to rederive them if you forget (say) the formula for the damped oscillator frequency or the criterion for critical damping.

As has been our general rule above, I think that it is better to *try* to make it through the derivation to where you understand it, even if only a single time and for a moment of understanding, even if you do then move on and just learn and work to retain the result. I think it helps you remember the result with less effort and for longer once the course is over, and to bring it back into mind and understand it more easily if you should ever need to in the future. But I also realize that mastering a chunk of math like this doesn't come easily to some of you and that investing the time to do *given* a limited amount of time to invest might actually reduce your eventual understanding of the general content of this chapter. You'll have to decide for yourself if this is true, ideally after at least giving the math below a look and a try. It's not really as difficult as it looks at first.

The exponential assumption:

$$x(t) = Ae^{\alpha t} \tag{9.16}$$

makes solutions to general linear homogeneous ODEs *simple*.

Let's look at the pattern when we take repeated derivatives of this equation:

$$\begin{aligned}
x(t) &= Ae^{\alpha t} \\
\frac{dx}{dt} &= \alpha Ae^{\alpha t} \\
\frac{d^2x}{dt^2} &= \alpha^2 Ae^{\alpha t} \\
\frac{d^3x}{dt^3} &= \alpha^3 Ae^{\alpha t} \\
\ldots
\end{aligned} \tag{9.17}$$

where α is an unknown parameter and A is an arbitrary (possibly complex) constant (with the units of $x(t)$, in this case, length) called the ***amplitude***. Indeed, this is a general rule:

$$\frac{d^nx}{dt^n} = \alpha^n Ae^{\alpha t} \tag{9.18}$$

for any $n = 0, 1, 2....$

Substituting this assumed solution and its rule for the second derivative into the SHO ODE, we get:

$$\begin{aligned}
\frac{d^2x}{dt^2} + \omega^2 x &= 0 \\
\frac{d^2 Ae^{\alpha t}}{dt^2} + \omega^2 Ae^{\alpha t} &= 0 \\
\alpha^2 Ae^{\alpha t} + \omega^2 Ae^{\alpha t} &= 0 \\
\left(\alpha^2 + \omega^2\right) Ae^{\alpha t} &= 0
\end{aligned} \tag{9.19}$$

There are two ways this equation could be true. First, we could have $A = 0$, in which case $x(t) = 0$ for any value of α. This indeed does solve the ODE, but the solution is *boring* – nothing moves! Mathematicians call this the ***trivial solution*** to a homogeneous linear ODE, and we will reject it out of hand by insisting that we have a *non*trivial solution with $A \neq 0$.

In that case it is necessary for

$$\left(\alpha^2 + \omega^2\right) = 0 \tag{9.20}$$

This is called the ***characteristic equation*** for the linear homogeneous ordinary differential equation. If we can find an α such that this equation is satisfied, then our assumed answer will indeed solve the ODE for nontrivial (nonzero) $x(t)$.

Clearly:

$$\alpha = \pm i\omega \tag{9.21}$$

where

$$i = +\sqrt{-1} \tag{9.22}$$

We now have a solution to our second order ODE – indeed, we have two solutions – but those solutions are ***complex exponentials***[212] and contain the ***imaginary unit***[213], i.

In principle, if you have satisfied the prerequisites for this course you have almost certainly studied imaginary numbers and complex numbers[214] in a high school algebra class and perhaps again in college level calculus. Unfortunately, because high school math is often indifferently well taught, you may have thought they would never be *good* for anything and hence didn't pay much attention to them, or (however well they were or were not covered) at this point you've now forgotten them.

In any of these cases, now might be a really ***good time to click on over to my online Mathematics for Introductory Physics book***[215] and review at least some of the properties of i and complex numbers and how they relate to trig functions. This book is still (as of this moment) less detailed here than I would like, but it does review all of their most important properties that are used below. Don't hesitate to follow the wikipedia links as well.

If you are a life science student (perhaps a bio major or premed) then *perhaps* (as noted above) you won't ever need to know even this much and can get away with just memorizing the real solutions below. If you are a physics or math major or an engineering student, the mathematics of this solution is ***just a starting point*** to an entire, amazing world of complex

[212] Wikipedia: http://www.wikipedia.org/wiki/Euler Formula.

[213] Wikipedia: http://www.wikipedia.org/wiki/Imaginary unit.

[214] Wikipedia: http://www.wikipedia.org/wiki/Complex numbers.

[215] http://www.phy.duke.edu/~rgb/Class/math_for_intro_physics.php There is an entire chapter on this: *Complex Numbers and Harmonic Trigonometric Functions*, well worth a look.

numbers, quaternions, Clifford (geometric division) algebras, that are not only *useful*, but seem to be *essential* in the development of electromagnetic and other field theories, theories of oscillations and waves, and above all in quantum theory (bearing in mind that everything we are learning this year is *technically* incorrect, because the Universe turns out not to be microscopically classical). Complex numbers also form the basis for one of the most powerful methods of doing certain classes of otherwise enormously difficult integrals in mathematics. So you'll have to decide for yourself just how far you want to pursue the discovery of this beautiful mathematics at this time – we will be presenting only the bare minimum necessary to obtain the desired, general, *real* solutions to the SHO ODE below.

Here are the two ***linearly independent solutions***:

$$x_+(t) = A_+ e^{+i\omega t} \tag{9.23}$$

$$x_-(t) = A_- e^{-i\omega t} \tag{9.24}$$

that follow, one for each possible value of α. Note that we will always get n independent solutions for an nth order linear ODE, because we will always have to solve for the roots of an nth order characteristic equation, and there are n of them! $A_\pm$ are the *complex* constants of integration – since the solution is complex already we might as well construct a general complex solution instead of a less general one where the $A_\pm$ are real.

Given these two independent solutions, an ***arbitrary, completely general*** solution can be made up of a sum of these two independent solutions:

$$x(t) = A_+ e^{+i\omega t} + A_- e^{-i\omega t} \tag{9.25}$$

We now use a pair of True Facts (that you can read about and see proven in the wikipedia articles linked above or in the online math review). First, let us note the *Euler Equation*:

$$e^{i\theta} = \cos(\theta) + i\sin(\theta) \tag{9.26}$$

This can be proven a number of ways – probably the easiest way to verify it is by noting the equality of the taylor series expansions of both sides – but we can just take it as "given" from here on in this class. Next let us note that a completely general complex number z can always be written as:

$$\begin{aligned} z &= x + iy \\ &= |z|\cos(\theta) + i|z|\sin(\theta) \\ &= |z|(\cos(\theta) + i\sin(\theta)) \\ &= |z|e^{i\theta} \end{aligned} \tag{9.27}$$

(where we used the Euler equation in the last step so that (for example) we can quite generally write:

$$A_+ = |A_+|e^{+i\phi_+} \tag{9.28}$$

$$A_- = |A_-|e^{-i\phi_-} \tag{9.29}$$

for some ***real*** amplitude $|A_\pm|$ and ***real*** phase angles $\pm\phi_\pm$[216].

[216] It doesn't matter if we define A_- with a negative phase angle, since ϕ_- might be a positive *or* negative number anyway. It can also always be reduced via modulus 2π into the interval $[0, 2\pi)$, because $e^{i\phi}$ is periodic.

If we substitute this into the two independent solutions above, we note that they can be written as:

$$x_+(t) = A_+ e^{+i\omega t} = |A_+|e^{i\phi_+}e^{+i\omega t} = |A_+|e^{+i(\omega t+\phi_+)} \tag{9.30}$$
$$x_-(t) = A_- e^{-i\omega t} = |A_-|e^{-i\phi_-}e^{-i\omega t} = |A_-|e^{-i(\omega t+\phi_-)} \tag{9.31}$$

Finally, we wake up from our mathematical daze, hypnotized by the strange beauty of all of these equations, smack ourselves on the forehead and say "But what am I *thinking!* I need $x(t)$ to be ***real***, because the physical mass m ***cannot possibly be found at an imaginary (or general complex) position***!". So we *take the real part of either of these solutions:*

$$\begin{aligned}\Re x_+(t) &= \Re|A_+|e^{+i(\omega t+\phi_+)} \\ &= |A_+|\Re\left(\cos(\omega t+\phi_+) + i\sin(\omega t+\phi_+)\right) \\ &= |A_+|\cos(\omega t+\phi_+)\end{aligned} \tag{9.32}$$

and

$$\begin{aligned}\Re x_-(t) &= \Re|A_-|e^{-i(\omega t+\phi_-)} \\ &= |A_-|\Re\left(\cos(\omega t+\phi_-) - i\sin(\omega t+\phi_-)\right) \\ &= |A_-|\cos(\omega t+\phi_-)\end{aligned} \tag{9.33}$$

These two solutions are the *same*. They differ in the (sign of the) *imaginary* part, but have exactly the same form for the real part. We have to figure out the amplitude and phase of the solution in any event (see below) and we won't get a *different* solution if we use $x_+(t)$, $x_-(t)$, or any linear combination of the two! We can finally get rid of the $\pm$ notation and with it, the last vestige of the complex solutions we used as an intermediary to get this lovely real solution to the position of (e.g.) the mass m as it oscillates connected to the perfect spring.

If you skipped ahead above, resume reading/studying here!

Thus:

$$x(t) = A\cos(\omega t+\phi) \tag{9.34}$$

is the ***completely general, real*** solution to the SHO ODE of motion, equation 9.14 above, valid in *any* context, including ones with a different context (and even a different variable) leading to a different algebraic form for ω^2.

A few final notes before we go on to try to *understand* this solution. There are two unknown real numbers in this solution, A and ϕ. These are the ***constants of integration!*** Although we didn't exactly "integrate" in the normal sense, we are still picking out a particular solution from an infinity of two-parameter solutions with different ***initial conditions***, just as we did for constant acceleration problems eight or nine weeks ago! If you like, this solution has to be able to describe the answer for *any permissible value of the initial position and velocity* of the mass at time $t = 0$. Since we can *independently* vary $x(0)$ and $v(0)$, we must have at least a two parameter *family* of solutions to be able to describe a general solution[217].

[217]In future courses, math or physics majors might have to cope with situations where you are given two pieces

9.1.2: The Simple Harmonic Oscillator Solution

As we *formally derived above*, the solution to the SHO equation of motion is;

$$x(t) = A\cos(\omega t + \phi) \tag{9.35}$$

where A is called the ***amplitude*** of the oscillation and ϕ is called the ***phase*** of the oscillation. The amplitude tells you how big the oscillation is at peak (maximum displacement from equilibrium); the phase tells you *when* the oscillator was started relative to your clock (the one that reads t). The amplitude has to have the same units as the variable, as $\sin, \cos, \tan, \exp$ functions (and their arguments) are all necessarily ***dimensionless*** in physics[218]. Note that we could have used $\sin(\omega t + \phi)$ as well, or any of several other forms, since $\cos(\theta) = \sin(\theta + \pi/2)$. But you knew that[219].

A and ϕ are two unknowns and have to be determined from the initial conditions, the givens of the problem, as noted above. They are basically constants of integration *just like* x_0 and v_0 for the one-dimensional constant acceleration problem. From this we can easily see that:

$$v(t) = \frac{dx}{dt} = -\omega A\sin(\omega t + \phi) \tag{9.36}$$

and

$$a(t) = \frac{d^2x}{dt^2} = -\omega^2 A\cos(\omega t + \phi) = -\frac{k}{m}x(t) \tag{9.37}$$

This last result ***proves that $x(t)$ solves the original differential equation*** and is where we would have gotten *directly* if we'd assumed a general cosine or sine solution instead of an exponential solution at the beginning of the previous section.

Note Well!

An ***unfortunately commonly made mistake*** for SHO problems is for students to take $F_x = ma = -kx$, write it as:

$$a = -\frac{k}{m}x \tag{9.38}$$

and then try to substitute this into the *kinematic* solutions for constant acceleration problems that we tried very hard not to blindly memorize back in weeks 1 and 2. That is, they try to write (for example):

$$x(t) = \frac{1}{2}at^2 + v_0 t + x_0 = -\frac{1}{2}\frac{k}{m}xt^2 + v_0 t + x_0 \tag{9.39}$$

This solution is ***so very, very wrong***, so wrong that it is ***deeply disturbing*** when students write it, as it means that they have *completely failed* to understand either the SHO or how to solve even the constant acceleration problem. Obviously it ***bears no resemblance*** to either

of *data* about the solution, not necessarily initial conditions. For example, you might be given $x(t_1)$ and $x(t_2)$ for two specified times t_1 and t_2 and be required to find the particular solution that satisfied these as a constraint. However, this problem is much more difficult and can easily be insufficient data to fully specify the solution to the problem. We will avoid it here and stick with initial value problems.

[218] All function in physics that have a power series expansion have to be dimensionless because we do not know how to add a liter to a meter, so to speak, or more generally how to add powers of any dimensioned unit.

[219] I hope. If not, time to review the unit circle and all those trig identities from 11th grade...

the correct answer or the observed behavior of a mass on a spring, which is to *oscillate*, not speed up quadratically in time. The appearance of x on both sides of the equation means that it isn't even a solution.

What it reveals is a student who has tried to learn physics by memorization, not by understanding, and hasn't even succeeded in that. Very sad.

Please do not make this mistake!

9.1.3: Plotting the Solution: Relations Involving ω

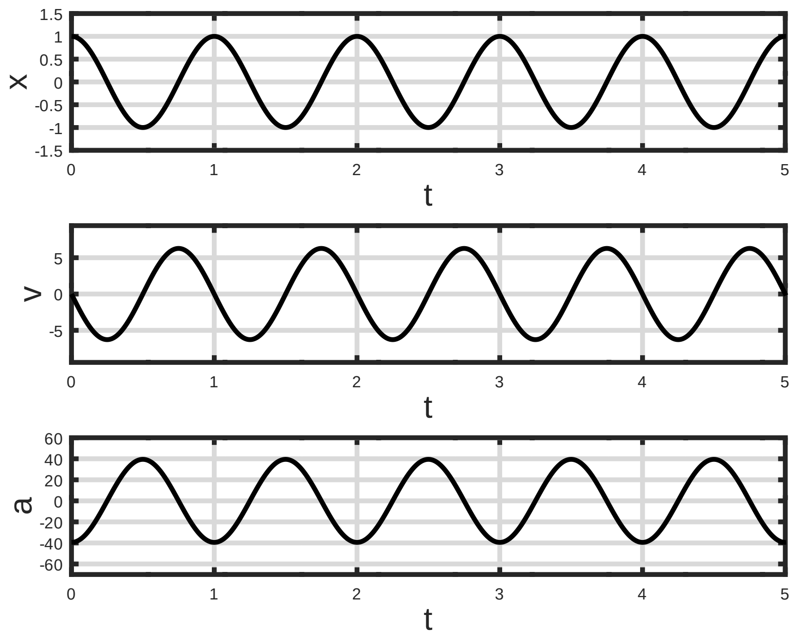


Figure 9.2: Solutions for a mass on a spring started at $x(0) = A = 1$, $v(0) = 0$ at time $t = 0$ (so that $\phi = 0$). Note well the location of the ***period of oscillation,*** $\boldsymbol{T = 1}$ on the time axis, one full cycle from the beginning.

Since we are going to quite a bit with harmonic oscillators from now on, we should take a few moments to ***plot*** $x(t)$, $v(t)$, and $a(t)$.

We remarked above that $omega$ had to have units of t^{-1}. The following are some True Facts involving ω that You Should Know:

$$\omega = \frac{2\pi}{T} \tag{9.40}$$

$$= 2\pi f \tag{9.41}$$

where T is the *period* of the oscillator the time required for it to return to an identical position *and* velocity) and f is called the *frequency* of the oscillator. Know these relations *instantly*.

They are easy to figure out but will cost you valuable time on a quiz or exam if you don't just take the time to completely embrace them now.

Note a very interesting thing. If we build a perfect simple harmonic oscillator, it oscillates at the same frequency *independent of its amplitude*. If we know the period and can count, we have just invented the *clock*. In fact, clocks are nearly *always* made out of various oscillators (why?); some of the earliest clocks were made using a pendulum as an oscillator and mechanical gears to count the oscillations, although now we use the much more precise oscillations of a bit of stressed crystalline quartz (for example) and electronic counters. The idea, however, remains the same.

9.1.4: Finding A and ϕ from Initial Values

Suppose a mass m is connected to a spring with spring constant k, is prepared and released in such a way that at $t = 0$, $x = X_0$ and $v = V_0$. To find A and ϕ, we write the equations representing $t = 0$ as:

$$x(0) = X_0 = A\cos(\omega \times 0 + \phi) \quad \Rightarrow$$
$$A\cos\phi = X_0 \tag{9.42}$$
$$v(0) = V_0 = -\omega A\sin(\omega \times 0 + \phi) \quad \Rightarrow$$
$$A\sin\phi = -V_0/\omega \tag{9.43}$$

We deliberately divide out the $-\omega$ from V_0 so that the latter equation has the same units and form as the first equation for as X_0. Next we ***divide these two equations*** to eliminate A and obtain an equation for ϕ:

$$\frac{\cancel{A}\sin\phi = -V_0/\omega}{\cancel{A}\cos\phi = X_0} \quad \Rightarrow \quad \tan\phi = -\frac{V_0}{\omega X_0} \quad \Rightarrow \quad \phi = \tan^{-1}\left(-\frac{V_0}{\omega X_0}\right) \tag{9.44}$$

Be careful with this result! The tangent function is periodic on $\boldsymbol{\pi}$, not $\boldsymbol{2\pi}$! This means that there are always *two* angles in the range $\phi \in [0, 2\pi)$ that solve this equation, in opposing quadrants on the unit circle. This choice is connected to the similar choice we discover next when solving for A.

To find A, we square and sum the two equations above:

$$A^2\cancel{(\cos^2\phi + \sin^2\phi)} = A^2 = \left(X_0^2 + \frac{V_0^2}{\omega^2}\right) \quad \Rightarrow \quad A = \pm\sqrt{\frac{\omega^2 X_0^2 + V_0^2}{\omega^2}} \tag{9.45}$$

(using the identity $\cos^2\phi + \sin^2\phi = 1$). Here you can see the problem repeated. Which sign should we use for A? Clearly choosing the positive sign is equivalent to using ϕ from one quadrant, but one can freely use the minus sign as long as we use ϕ from the *other* quadrant! There are several ways to resolve the conflict, but my own preference is to (say) always choose solutions to the inverse tangent in the range $\phi \in [0, \pi)$, and the select A with the sign that then empirically agrees with the initial conditions, but (sigh) hopefully your grader will be sufficiently familiar with the other equivalent representation to avoid taking points off for a correct answer.

As usual, I strongly recommend against *memorizing* these results – the *method by which they are obtained* is far more useful, as it might be used if you were given data such as

$x(10) = X_{10}$, $v(4) = V_4$, or $x(4) = X_4$, $x(10) = X_{10}$. In other words one can use (almost) *any* two conditions, including ones given at *two different times* to obtail the constants of integration by simply substituting in the times and working to solve for the constants.

While you *should* be able do the algebra on the previous page if quizzed or tested with a problem involving ***general*** X_0 and V_0, the most *likely* initial conditions you will encounter are drawn from a much smaller set. You should be able to do the following ***four*** special cases without resorting to the algebra above:

$$x(0) = \pm X_0, v(0) = 0 \tag{9.46}$$

or

$$x(0) = 0, v(0) = \pm V_0 \tag{9.47}$$

Be sure you understand what these initial conditions represent. The first two represent a mass ***released from rest*** from some initial non-equilibrium position on the (stretched $+X_0$ or compressed $-X_0$) spring. These are the most likely initial conditions you will encounter in problems and examples, as they are the conditions generally used when one is taught or tested on only a general knowledge of "how harmonic oscillators work". The second set correspond to being given an initial velocity (to the right $+V_0$ or left $-V_0$) *starting at the equilibrium position*, $x(0) = 0$. Problems of this sort often involve some sort of "collision" with a mass attached to a spring and initially sitting at rest at equilibrium.

Here are the solutions corresponding to these "special" initial conditions. ***Know these without all the algebra!*** In the first case, $\boldsymbol{\phi = 0}$ and $\boldsymbol{A = \pm X_0}$ so that:

$$x(t) = \pm X_0 \cos(\omega t) \quad \Rightarrow \quad v(t) = \mp \omega X_0 \sin(\omega t) \tag{9.48}$$

Note well that this implements the suggestion above. I've picked $\phi = 0$ (allowing the phase to be omitted entirely from the solutions) and allowed the amplitude to carry the sign of the initial condition X_0.

In the second case, the solutions we might obtain from the inverse tangent are $\phi = \pm\pi/2$. However, note well the trigonometric identity: $\cos(\theta \pm \pi/2) = \mp \sin(\theta)$! This gives us *yet another* way of representing a solution! It turns out to be a ***lot easier*** to just write the solutions reversing the role of $\sin$ and $\cos$, with no phase angle required once again! That is:

$$v(t) = \pm V_0 \cos(\omega t) = \pm \omega A \cos(\omega t) \quad \Rightarrow \quad A = \frac{V_0}{\omega} \quad \Rightarrow \quad x(t) = \pm \frac{V_0}{\omega} \sin(\omega t) \tag{9.49}$$

It's easy to check these using $\dfrac{d}{dt} \cos \omega t = -\omega \sin \omega t$ and $\dfrac{d}{dt} \sin \omega t = \omega \cos \omega t$.

Example 9.1.1: A Mass on a Spring Driven to the Left

Suppose a mass m is attached to a spring with spring constant k and left at equilibrium on a frictionless surface as displayed in figure 9.3. At $t = 0$, it is struck with a sudden blow, driving it with the ***impulse*** "instantly" to the left at initial speed v_0. Find $x(t)$ and $v(t)$!

Solution: I'll skip the details of actually solving the differential equation of motion:

$$F_x = -kx = \frac{d^2x}{dt^2} \quad \Rightarrow \quad \frac{d^2x}{dt^2} + \frac{k}{m} x = 0 \quad \Rightarrow \quad \omega = \sqrt{\frac{k}{m}} \tag{9.50}$$

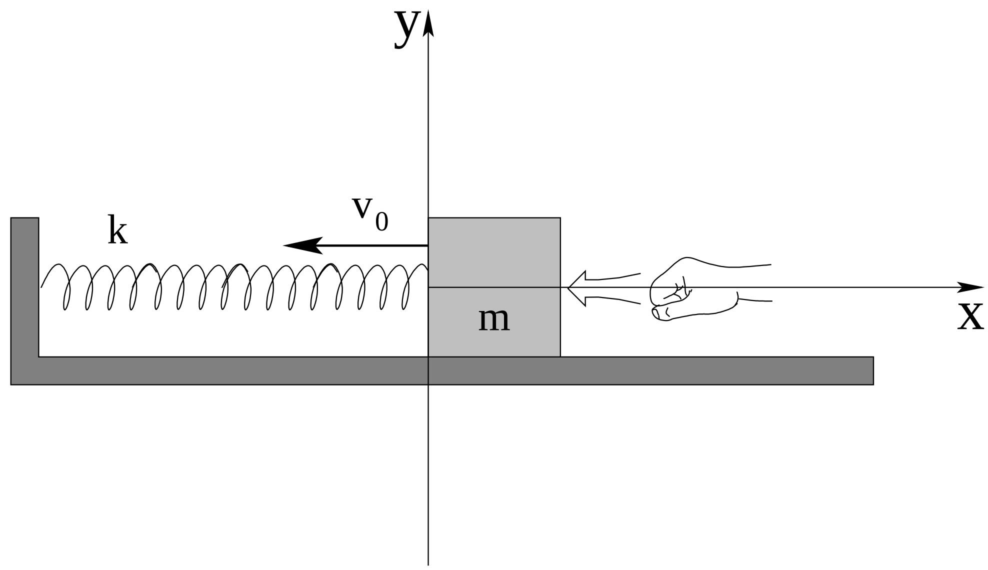


Figure 9.3: A mass on a spring at equilibrium is struck a sudden blow at $t = 0$ to give it an initial velocity of $-v_0$, to the *left*.

and note that this is precisely one of the four "special" initial conditions you should recognize. It is trivial to get:

$$v(t) = -v_0 \cos(\omega t) \tag{9.51}$$

which clearly satisfies $v(0) = -v_0$. We know that (corresponding to this):

$$x(t) = A\sin(\omega t) \quad \Rightarrow \quad A = -\frac{v_0}{\omega} \quad \Rightarrow \quad x(t) = -\frac{v_0}{\omega}\sin(\omega t) \tag{9.52}$$

Finally, we should assemble a finl statement of the solution *in terms of the givens* as best practice, although the solution above is already clear:

$$\boxed{x(t) = -\sqrt{\frac{m}{k}} v_0 \sin\left(\sqrt{\frac{k}{m}} t\right)} \tag{9.53}$$

$$\boxed{v(t) = -v_0 \cos\left(\sqrt{\frac{k}{m}} t\right)} \tag{9.54}$$

9.1.5: The Energy of a Mass on a Spring

As we evaluated and discussed in week 3, the spring exerts a ***conservative force*** on the mass m. Thus:

$$\begin{aligned} U &= -W(0 \to x) = -\int_0^x (-kx)dx = \frac{1}{2}kx^2 \\ &= \frac{1}{2}kA^2\cos^2(\omega t + \phi) \end{aligned} \tag{9.55}$$

where we have arbitrarily set the zero of potential energy and the zero of the coordinate system to be the equilibrium position[220].

[220] What would it look like if the zero of the energy were at an arbitrary $x = x_0$? What would the force and energy look like if the zero of the coordinates where at the point where the spring is attached to the wall?

The kinetic energy is:

$$\begin{aligned} K &= \frac{1}{2}mv^2 \\ &= \frac{1}{2}m(\omega^2)A^2\sin^2(\omega t+\phi) \\ &= \frac{1}{2}m(\frac{k}{m})A^2\sin^2(\omega t+\phi) \\ &= \frac{1}{2}kA^2\sin^2(\omega t+\phi) \end{aligned} \tag{9.56}$$

The total energy is thus:

$$\begin{aligned} E &= \frac{1}{2}kA^2\sin^2(\omega t+\phi)+\frac{1}{2}kA^2\cos^2(\omega t+\phi) \\ &= \frac{1}{2}kA^2 \end{aligned} \tag{9.57}$$

and is *constant* in time! Kinda spooky how that works out...

Note that the energy oscillates between being all potential at the extreme ends of the swing (where the object comes to rest) and all kinetic at the equilibrium position (where the object experiences no force).

9.1.6: Mass Hanging on a Spring

One problem with our original harmonic oscillator model is that it is *very* difficult to arrange a mass on a spring that slides back and forth on a *table* (to oppose gravity with a normal force) without introducing an unacceptable non-conservative *force of kinetic friction that opposes the sliding*. This greatly complicates the solution, especially since that force of kinetic friction is not likely to be *small*, and difficult to *make* small enough to actually be ignored, short of putting the mass on an air track of some sort!

On the other hand, it is *easy* to just *hang a mass from a vertical spring*. If you do this, and pull down the hanging mass, it *does* indeed oscillate, and the only non-conservative damping force is *air drag* (plus a teensy loss heating the spring itself as it bends), so one can at least *hope* that neglecting drag will yield solutions that correspond well with observations, or that we might even be able to treat drag explicitly!

But ***what about gravity?*** Specifically, will it *screw up the equation of motion* so it is no longer solvable without great effort?

In figure 9.4 (first panel) a mass m is shown attached to a vertical spring with spring constant k. Note well that I've drawn coordinates in such that y is positive *down* (in the mg direction).

It is gently lowered to a *new* equilibrium postion y_0 (second panel), where

$$mg - ky_0 = 0 \quad\Rightarrow\quad y_0 = \frac{mg}{k} \tag{9.58}$$

and it remains at rest.

It is then stretched by an *additional* distance y' (third panel) so that:

$$F_y = mg - k(y_0 + y') = \cancel{(mg - ky_0)} - ky' = -ky' = m\frac{d^2y}{dt^2} \tag{9.59}$$

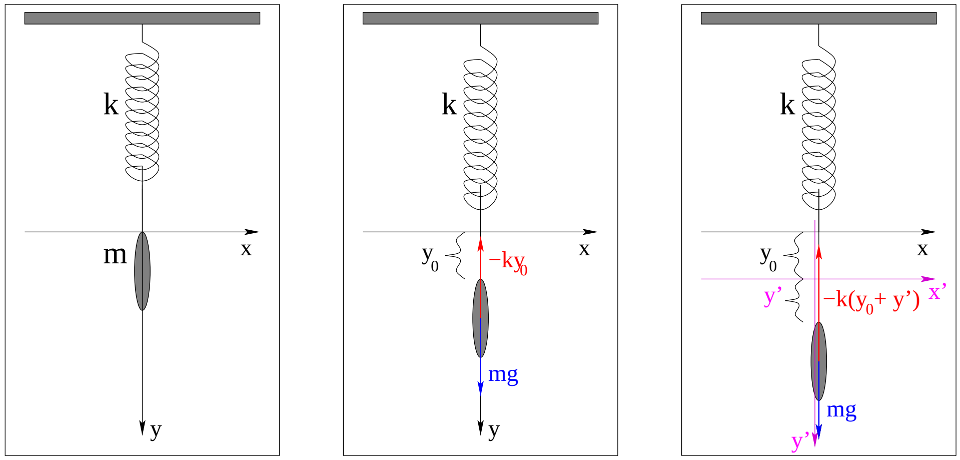


Figure 9.4: A streamlined mass attached to a vertical spring as it is stretched first to the new equilbrium at y_0, then pulled down still further by a distance y'.

Note that the mg term explicitly cancels *part* of the spring force!

We then ***change variables to*** $\boldsymbol{y' = y - y_0}$, basically measuring y' from the magenta axis added to the third panel. Note that:

$$\frac{d^2y}{dt^2} = \frac{d^2y'}{dt^2}$$

because y_0 is a *constant!*

N2 and the equation of motion in the y' coordinates then become:

$$F_{y'} = -ky' = m\frac{d^2y'}{dt^2} \quad \Rightarrow \quad \boxed{\frac{d^2y'}{dt^2} + \frac{k}{m}y' = 0} \tag{9.60}$$

Which is the SHOE with:

$$\omega = \sqrt{\frac{k}{m}}$$

as usual!

We conclude that a mass hanging on a spring ***still oscillates harmonically around the new equilibrium*** y_0 with its frequency ω etc ***unchanged***. Specifically and especially, *no* g *appears in the solution* in y' coordinates, and we can always recover a solution in the original coordinates as e.g.

$$y'(t) = A\cos(\omega t + \phi) \quad \Rightarrow \quad y(t) = \frac{mg}{k} + A\cos(\omega t + \phi) = y_0 + y'(x,t) \tag{9.61}$$

Now all that is damping our streamlined mass is whatever fluid it is moving through. We can hang the whole apparatus in a vacuum chamber, and eliminate even that! At that point the *only* dissipative force is going to be the tiny inelastic forces within the spring itself as it flexes, that will *very, very slowly* convert the original oscillatory energy into heat, damping out the amplitude.

In a little bit, we'll use this arrangement to *add a linear damping force in a controllable way* and end up with an equation of motion for a ***linearly damped simple harmonic oscillator*** that we will actually be able to solve! That solution (simple as it is) will turn out to be enormously useful all the way through graduate physics courses in a variety of contexts.

This more or less concludes our *general* discussion of simple harmonic oscillators in the specific context of a mass on a spring, bearing in mind that we have yet to treat damping, harmonic driving forces, and resonance. However, there are many more systems that oscillate harmonically, or nearly harmonically that go far beyond “just” actual masses on actual springs. Let’s study another very important one next.

9.2: The Pendulum

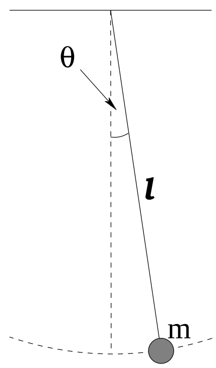


Figure 9.5: A simple pendulum is just a point like mass suspended on a long string and displaced sideways by a small angle. We will assume no damping forces and that there is no initial velocity into or out of the page, so that the motion is stricly in the plane of the page.

The pendulum is another example of a simple harmonic oscillator, at least for small oscillations. Suppose we have a mass m attached to a string of length ℓ. We swing it up so that the stretched string makes a (small) angle θ_0 with the vertical and release it at some time (not necessarily $t = 0$). What happens?

We write Newton’s Second Law for the force component *tangent* to the arc of the circle of the swing as:

$$F_t = -mg\sin(\theta) = ma_t = m\ell\frac{d^2\theta}{dt^2} \tag{9.62}$$

where the latter follows from $a_t = \ell\alpha$ (the angular acceleration). Then we rearrange to get:

$$\frac{d^2\theta}{dt^2} + \frac{g}{\ell}\sin(\theta) = 0 \tag{9.63}$$

This is *almost* a simple harmonic equation. To make it one, we have to use the small angle approximation:

$$\sin(\theta) \approx \theta \tag{9.64}$$

Then:

$$\frac{d^2\theta}{dt^2} + \frac{g}{\ell}\theta = \frac{d^2\theta}{dt^2} + \omega^2\theta = 0 \tag{9.65}$$

where we have defined

$$\omega^2 = \frac{g}{\ell} \tag{9.66}$$

and we can just ***write down the solution***:

$$\theta(t) = \Theta\cos(\omega t + \phi) \tag{9.67}$$

with $\omega = \sqrt{\frac{g}{\ell}}$, Θ the amplitude of the oscillation, and phase ϕ just as before.

Now you see the advantage of all of our hard work in the last section. To solve ***any SHO problem*** one simply puts the differential equation of motion (approximating as necessary) into the form of the SHO ODE ***which we have solved once and for all above!*** We can then just write down the solution and be quite confident that all of its features will be "just like" the features of the solution for a mass on a spring.

For example, if you compute the gravitational potential energy for the pendulum for *arbitrary* angle θ, you get:

$$U(\theta) = mg\ell\,(1 - \cos(\theta)) \tag{9.68}$$

This doesn't initially *look* like the form we might expect from blindly substituting similar terms into the potential energy for mass on the spring, $U(t) = \frac{1}{2}kx(t)^2$. "$k$" for the gravity problem is $m\omega^2$, "$x(t)$" is $\theta(t)$, so:

$$U(t) = \frac{1}{2}mg\ell\Theta^2 \sin^2(\omega t + \phi) \tag{9.69}$$

is what we expect.

As an interesting and fun exercise (that really isn't too difficult) see if you can prove that these two forms are really the same, *if* you make the small angle approximation $\theta \ll 1$ in the first form! This shows you pretty much where the approximation will *break down* as Θ is gradually increased. For large enough θ, the period of a pendulum clock *does* depend on the amplitude of the swing. This (and the next section) explains grandfather clocks – clocks with very long penduli that can swing very slowly through very small angles – and why they were so accurate for their day.

9.2.1: The Physical Pendulum

In the treatment of the ordinary pendulum above, we just used Newton's Second Law directly to get the equation of motion. This was possible only because we could neglect the mass of the string and because we could treat the mass like a point mass at its end, so that its moment of inertia was (if you like) just $m\ell^2$.

That is, we *could* have solved it using Newton's Second Law for *rotation* instead. If θ in figure 9.5 is positive (out of the page), then the torque due to gravity is:

$$\tau = -mg\ell\sin(\theta) \tag{9.70}$$

and we can get to the *same equation of motion* via:

$$\begin{aligned} I\alpha \;=\; m\ell^2\frac{d^2\theta}{dt^2} &= -mg\ell\sin(\theta) \;=\; \tau \\ \frac{d^2\theta}{dt^2} &= -\frac{g}{\ell}\sin(\theta) \\ \frac{d^2\theta}{dt^2} + \frac{g}{\ell}\sin(\theta) &= 0 \\ \frac{d^2\theta}{dt^2} + \omega^2\theta \;=\; \frac{d^2\theta}{dt^2} + \frac{g}{\ell}\theta &= 0 \end{aligned} \tag{9.71}$$

(where we use the small angle approximation in the last step as before).

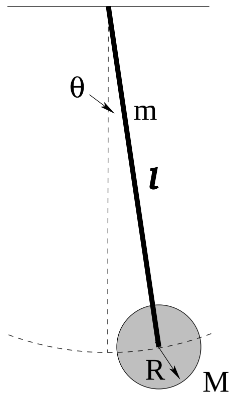


Figure 9.6: A physical pendulum takes into account the fact that the masses that make up the pendulum have a total ***moment of inertia*** about the pivot of the pendulum.

However, *real* grandfather clocks often have a large, massive pendulum like the one above pictured in figure 9.6 – a long massive rod (of length ℓ and uniform mass m) with a large round disk (of radius R and mass M) at the end. Both the rod and disk rotate about the pivot with each oscillation; they have *angular momemtum*. Newton's Law for forces alone no longer suffices. We must use torque and the moment of inertia (found using the parallel axis theorem) to obtain the frequency of the oscillator[221].

To do this we go through the *same steps* that I just did for the simple pendulum. The only real difference is that now the weight of both masses contribute to the torque (and the force exerted by the pivot can be ignored), and as noted we have to work harder to compute the moment of inertia.

So let's start by computing the net gravitational torque on the system at an arbitrary (small) angle θ. We get a contribution from the rod (where the weight acts "at the center of mass" of the rod) and from the pendulum disk:

$$\tau = -\left(mg\frac{\ell}{2} + Mg\ell\right)\sin(\theta) \tag{9.72}$$

The negative sign is there because the torque *opposes* the angular displacement from equilibrium and points *into* the page as drawn.

Next we set this equal to $I\alpha$, where I is the total moment of inertia for the *system* about the pivot of the pendulum and simplify:

$$\begin{aligned}
I\alpha \;=\; I\frac{d^2\theta}{dt^2} &= -\left(mg\frac{\ell}{2} + Mg\ell\right)\sin(\theta) \;=\; \tau \\
\frac{d^2\theta}{dt^2} &= -\frac{\left(mg\frac{\ell}{2} + Mg\ell\right)}{I}\sin(\theta) \\
\frac{d^2\theta}{dt^2} + \frac{\left(mg\frac{\ell}{2} + Mg\ell\right)}{I}\sin(\theta) &= 0 \\
\frac{d^2\theta}{dt^2} + \frac{\left(mg\frac{\ell}{2} + Mg\ell\right)}{I}\theta &= 0
\end{aligned} \tag{9.73}$$

[221] I know, I know, you had hoped that you could finally forget all of that stressful stuff we learned back in the weeks we covered torque. Sorry. Not happening.

where we finish off with the small angle approximation as usual for pendulums. We can now recognize that this ODE has the standard form of the SHO ODE:

$$\frac{d^2\theta}{dt^2} + \omega^2\theta = 0 \tag{9.74}$$

with

$$\omega^2 = \frac{\left(mg\frac{\ell}{2} + Mg\ell\right)}{I} \tag{9.75}$$

I left the result in terms of I because it is simpler that way, but of course we have to evaluate I in order to evaluate ω^2. Using the parallel axis theorem (and/or the moment of inertia of a rod about a pivot through one end) we get:

$$I = \frac{1}{3}m\ell^2 + \frac{1}{2}MR^2 + M\ell^2 \tag{9.76}$$

This is "the moment of inertia of the rod plus the moment of inertia of the disk rotating about a parallel axis a distance ℓ away from its center of mass". From this we can read off the angular frequency:

$$\omega^2 = \frac{4\pi^2}{T^2} = \frac{\left(mg\frac{\ell}{2} + Mg\ell\right)}{\frac{1}{3}m\ell^2 + \frac{1}{2}MR^2 + M\ell^2} \tag{9.77}$$

With ω in hand, we know everything. For example:

$$\theta(t) = \Theta\cos(\omega t + \phi) \tag{9.78}$$

gives us the angular trajectory. We can easily solve for the period T, the frequency $f = 1/T$, the spatial or angular velocity, or whatever we like.

Note that the energy of this sort of pendulum can be tricky, not because it is conceptually any different from before but because there are so many symbols in the answer. For example, its potential energy is easy enough – it depends on the elevation of the center of masses of the rod and the disk. The

$$U(t) = (mgh(t) + MgH(t)) = \left(mg\frac{\ell}{2} + Mg\ell\right)(1 - \cos(\theta(t))) \tag{9.79}$$

where hopefully it is obvious that $h(t) = \ell/2\ (1-\cos(\theta(t)))$ and $H(t) = \ell\ (1-\cos(\theta(t))) = 2h(t)$. Note that the time dependence is entirely inherited from the fact that $\theta(t)$ is a function of time.

The kinetic energy is given by:

$$K(t) = \frac{1}{2}I\Omega^2 \tag{9.80}$$

where $\Omega = d\theta/dt$ as usual.

We can easily evaluate:

$$\Omega = \frac{d\theta}{dt} = -\omega\Theta\sin(\omega t + \phi) \tag{9.81}$$

so that

$$K(t) = \frac{1}{2}I\Omega^2 = \frac{1}{2}I\omega^2\Theta^2\sin^2(\omega t + \phi) \tag{9.82}$$

Recalling the definition of ω^2 above, this simplifies to:

$$K(t) = \frac{1}{2}\left(mg\frac{\ell}{2} + Mg\ell\right)\Theta^2\sin^2(\omega t + \phi) \tag{9.83}$$

so that:

$$\begin{aligned} E_{\rm tot} &= U + K \\ &= \left(mg\frac{\ell}{2} + Mg\ell\right)(1-\cos(\theta(t))) + \frac{1}{2}\left(mg\frac{\ell}{2} + Mg\ell\right)\Theta^2\sin^2(\omega t + \phi) \end{aligned} \quad (9.84)$$

which is not, in fact, a constant.

However, for small angles (the only situation where our solution is valid, actually) it is *approximately* a constant as we will now show. The trick is to use the Taylor series for the cosine function:

$$\cos(\theta) = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} + ... \quad (9.85)$$

and keep only the first term:

$$1 - \cos(\theta) = \frac{\theta^2}{2!} + \frac{\theta^4}{4!} + ... \approx \frac{\theta^2}{2} = \frac{1}{2}\Theta^2\cos^2(\omega t + \phi) \quad (9.86)$$

You should now be able to see that in fact, the total energy of the oscillator *is* "constant" in the small angle approximation.

Of course, it is actually constant even for *large* oscillations, but proving this requires solving the exact ODE with the $\sin(\theta)$ in it. This ODE is a version of the Sine-Gordon equation and has an elliptic integral for a solution that is way, way beyond the level of this course and indeed is right up there at the edge of some serious (but as always, way cool) math. We will stick with small angles!

9.3: The Torsional Oscillator

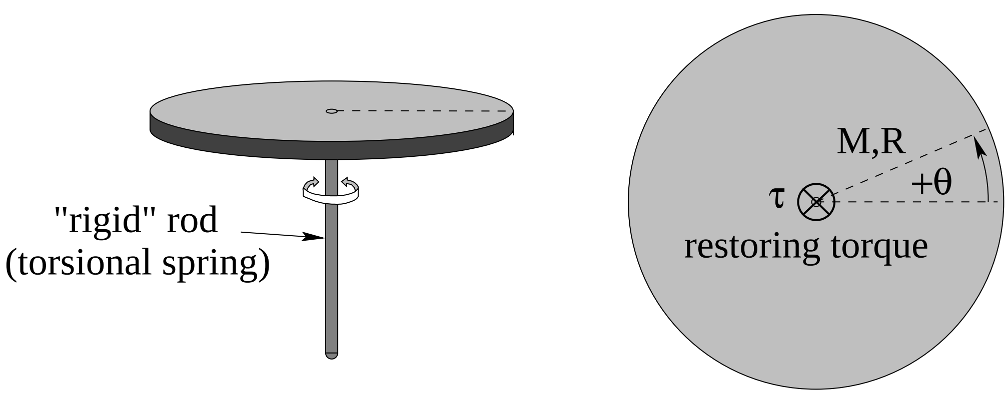


Figure 9.7: A torsional pendulum consist of an e.g. disk mounted on a "rigid" rod that behaves like a *torsional spring*, exerting a *restoring torque* proportional to the angle of rotation. If θ is positive out of the page, then $\vec{\tau}$ is *into* the page as shown above, *back* towards equilibrium.

Yet another example of an oscillator is pictured in figure 9.7 – a disk of mass M and radius R mounted on a rigid rod. When the disk is twisted, it imparts a *shear twist* onto the rod, which then attempts to twist back into its original untwisted shape. The rod, in fact, behaves like a *torsional spring*, exerting a *linear restoring torque* on the disk that tries to twist it back to its original equilibrium position:

$$\tau = -\kappa\theta \quad (9.87)$$

In this equation, κ is the "torsional spring constant" and plays exactly the same role as the usual spring constant k used above.

We can now easily write an equation of motion for the disk. If it is twisted to an initial angle θ_0 and released from rest, it will *harmonically oscillate!* At an arbitrary angle θ, the equation of motion is:

$$\tau = -\kappa\theta = I\alpha = \frac{1}{2}MR^2\frac{d^2\theta}{dt^2} \quad \Rightarrow \quad \boxed{\frac{d^2\theta}{dt^2} + \frac{2\kappa}{MR^2}\theta = 0} \tag{9.88}$$

which clearly has the standard form of the SHOE. We can "read off":

$$\omega = \sqrt{\frac{2\kappa}{MR^2}} \tag{9.89}$$

and write the solution:

$$\theta(t) = \theta_0 \cos\left(\sqrt{\frac{2\kappa}{MR^2}}t\right) \tag{9.90}$$

Note well that we *did not need* a small angle approximation here because the torsional spring is be approximately *linear* in the usual "elastic" regime for the rod, that is, for twists that are not so large as to risk deforming the rod.

Of course, many things can be varied for problems involving torsional oscillators. The torsional spring constant. The moment of inertia of the twisted mass connected to the torsional spring. The initial conditions. One can even be presented with a rotational collision problem where a mass is thrown to stick to the disk and thus impart its initial rotation, leading to oscillation! Be prepared!

9.4: Damped Oscillation

So far, all the oscillators we've treated are *ideal*. There is no friction or damping. In the real world, of course, things *always* damp down. You have to keep pushing the kid on the swing or they slowly come to rest. Your car doesn't *keep* bouncing after going through a pothole in the road. Buildings and bridges, clocks and kids, real oscillators all have damping.

Damping forces can be very complicated. There is kinetic friction, which tends to be independent of speed. There are various fluid drag forces, which tend to depend on speed, but in a sometimes complicated way depending on the shape of the object and e.g. the Reynolds number, as flow around it converts from laminar to turbulent. There may be other forces that we haven't studied yet that contribute at least weakly to damping[222]. So in order to get beyond a very qualitative description of damping, we're going to have to specify a *form* for the damping force (ideally one we can work with, i.e. integrate).

Note that we already showed in a previous section that we can hang a mass on a spring and measure y from its *new* equilibrium position and *still end up with the SHOE* in the new coordinate – gravity cancels *part* of the spring force, and the remainder is still linear in the

[222] Such as gravitational damping – an oscillating mass interacts with its massive environment to very, very slowly convert its organized energy into heat. We're talking slowly, mind you. Still, fast enough that the moon is gravitationally locked with the earth over geological times, and e.g. tidal/gravitational forces heat the moon Europa (as it orbits Jupiter) to the point where it is speculated that there is liquid water under the ice on its surface...

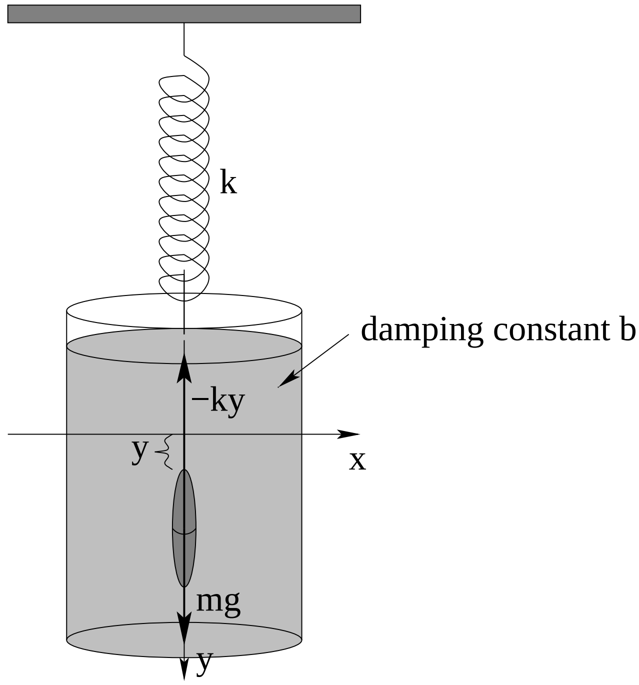


Figure 9.8: A smooth convex mass on a spring that is immersed in a suitable damping liquid experiences a *linear damping force* due to viscous interaction with the fluid in laminar flow. This idealizes the forces to where we can solve them and understand semi-quantitatively how to describe damped oscillation.

(new) displacement and back to equilibrium from both sides. This gives us a clear path to add a controlled damping force!

So that is exactly what we'll do! The simplest model of damped oscillation, illustrated in figure 9.8, produces a ***linear damping force***:

$$F_d = -bv \tag{9.91}$$

such as we would expect to observe in ***laminar flow*** around the *streamlined* oscillating object as long as it moves at speeds too low to excite turbulence in the surrounding fluid. (See e.g. Week 2 or 8 for further discussion of linear damping forces.)

With this form we can get an exact solution to the differential equation easily (good), get a preview of a solution we'll need next semester to study LRC circuits (better), and get a very nice qualitative picture of damping even when the damping force is *not* precisely linear (best). It will even prove relevant in electrodynamics where it and a suitable basis for molecular polarizability will lead us to a model for *dispersion!*

So let's do it! We start by writing Newton's Second Law for a mass m on a spring with spring constant k and a damping force $-bv$:

$$F_x = -kx - bv = ma = m\frac{d^2x}{dt^2} \tag{9.92}$$

Again, simple manipulation leads to:

$$\frac{d^2x}{dt^2} + \frac{b}{m}\frac{dx}{dt} + \frac{k}{m}x = 0 \tag{9.93}$$

which is the "standard form" for a damped mass on a spring and (within fairly obvious substitutions) for the general linearly damped SHO.

This is still a linear, second order, homogeneous, ordinary differential equation, but now we ***cannot just guess*** $\boldsymbol{x(t) = A\cos(\omega t)}$ because the first derivative of a cosine is a sine! This time we really ***must*** guess that $x(t)$ is a function that is proportional to its own first derivative!

We therefore guess $x(t) = Ae^{\alpha t}$ ***as before***, substitute for $x(t)$ and its derivatives, and get:

$$\left(\alpha^2 + \frac{b}{m}\alpha + \frac{k}{m}\right) Ae^{\alpha t} = 0 \tag{9.94}$$

As before, we exclude the trivial solution $x(t) = 0$ as being too boring to solve for (requiring that $A \neq 0$, that is) and are left with the characteristic equation for α:

$$\alpha^2 + \frac{b}{m}\alpha + \frac{k}{m} = 0 \tag{9.95}$$

This quadratic equation must be satisfied in order for our guess to be a nontrivial solution to the damped SHO ODE.

To solve for α we have to use the ***dread quadratic formula***:

$$\alpha = \frac{\frac{-b}{m} \pm \sqrt{\frac{b^2}{m^2} - \frac{4k}{m}}}{2} \tag{9.96}$$

This isn't quite where we want it. We expect from experience and intuition that for ***weak damping*** we should get an oscillating solution, indeed one that (in the limit that $b \to 0$) turns back into our familiar solution to the undamped SHO above. In order to get an oscillating solution, the argument of the square root must be ***negative*** so that our solution becomes a complex exponential solution as before!

This motivates us to factor a $-4k/m$ out from under the radical (where it becomes $i\omega_0$, where $\omega_0 = \sqrt{k/m}$ is the frequency of the *undamped* oscillator with the same mass and spring constant). In addition, we simplify the first term and get:

$$\alpha = \frac{-b}{2m} \pm i\omega_0\sqrt{1 - \frac{b^2}{4km}} \tag{9.97}$$

As was the case for the undamped SHO, there are two solutions:

$$x_{\pm}(t) = A_{\pm}e^{\frac{-b}{2m}t}e^{\pm i\omega' t} \tag{9.98}$$

where

$$\omega' = \omega_0\sqrt{1 - \frac{b^2}{4km}} \tag{9.99}$$

This, you will note, is ***not terribly simple or easy to remember!*** Yet you are responsible for knowing it. You have the usual choice – work very hard to memorize it, or learn to do the derivation(s).

I personally ***do not remember it at all*** save for a week or two around the time I teach it each semester. Too big of a pain, too easy to derive if I need it. But here you must suit yourself – either memorize it the same way that you'd memorize the digits of π, by lots and lots of mindless practice, or learn how to solve the equation, as you prefer.

Without recapitulating the entire argument, it should be fairly obvious that can take the real part of their sum, get formally identical terms, and combine them to get the general real solution:

$$x_{\pm}(t) = Ae^{\frac{-bt}{2m}} \cos(\omega' t + \phi) \tag{9.100}$$

where A is the real initial amplitude and ϕ determines the relative phase of the oscillator. The only two differences, then, are that the frequency of the oscillator is *shifted* to ω' and the whole solution is ***exponentially damped in time***.

9.4.1: Properties of the Damped Oscillator

There are several properties of the damped oscillator that are important to know.

- The amplitude damps *exponentially* as time advances. After a certain amount of time, the amplitude is halved. After the *same* amount of time, it is halved again.
- The frequency ω' is shifted so that it is ***smaller than*** ω_0, the frequency of the identical but undamped oscillator with the same mass and spring constant.
- The oscillator can be ***(under)damped***, ***critically damped***, or ***overdamped***. These terms are defined below.
- For exponential decay problems, recall that it is often convenient to define the ***exponential decay time***, in this case:

  $$\tau = \frac{2m}{b} \tag{9.101}$$

  This is the time required for the amplitude to go down to $1/e$ of its value from *any* starting time. For the purpose of drawing plots, you can imagine $e = 2.718281828 \approx 3$ so that $1/e \approx 1/3$. Pay attention to how the damping time ***scales*** with m and b. This will help you develop a conceptual understanding of damping.

Several of these properties are illustrated in figure 9.9. In this figure the ***exponential envelope*** of the damping is illustrated – this envelope determines the maximum amplitude of the oscillation as the total energy of the oscillating mass decays, turned into heat in the damping fluid. The period T' is indeed longer, but even for this relatively rapid damping, it is still nearly identical to T_0! See if you can determine what ω' is in terms of ω_0 numerically given that $\omega_0 = 2\pi$ and $b/m = 0.3$. Pretty close, right?

This oscillator is ***underdamped***. An oscillator is underdamped if ω' is ***real***, which will be true if:

$$\frac{b^2}{4m^2} = \left(\frac{b}{2m}\right)^2 < \frac{k}{m} = \omega_0^2 \tag{9.102}$$

An underdamped oscillator will exhibit true oscillations, eventually (exponentially) approaching zero amplitude due to damping.

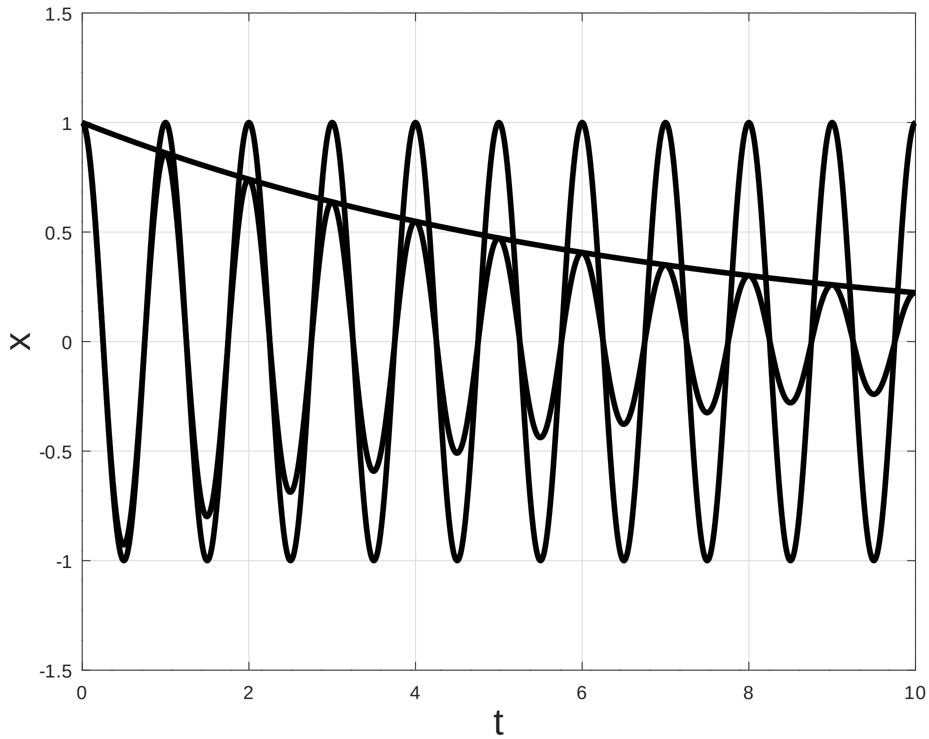


Figure 9.9: Two identical oscillators, one undamped (with $\omega_0 = 2\pi$, or if you prefer with an undamped period $T_0 = 1$) and one weakly damped ($b/m = 0.3$).

The oscillator is ***critically damped*** if ω' is ***zero***. This occurs when:

$$\frac{b^2}{4m^2} = \left(\frac{b}{2m}\right)^2 = \frac{k}{m} = \omega_0^2 \tag{9.103}$$

The oscillator will then *not oscillate* – it will go to zero *exponentially* in the shortest possible time. This (and barely underdamped and overdamped oscillators) is illustrated in figure 9.10.

The oscillator is *overdamped* if ω' is imaginary, which will be true if

$$\frac{b^2}{4m^2} = \left(\frac{b}{2m}\right)^2 > \frac{k}{m} = \omega_0^2 \tag{9.104}$$

In this case α ***is entirely real*** and has a component that damps very slowly. The amplitude goes to zero exponentially as before, but over a longer (possibly much longer) time and does not oscillate through zero at all.

Note that these inequalities and equalities that establish the ***critical boundary*** between oscillating and non-oscillating solutions involve the relative size of the inverse time constant associated with *damping* compared to the inverse time constant (times 2π) associated with *oscillation*. When the former (damping) is larger than the latter (oscillation), damping *wins* and the solution is *overdamped*. When it is smaller, oscillation wins and the solution is under-damped. When they are precisely equal, oscillation precisely disappears, k isn't *quite* strong enough compared to b to give the mass enough momentum to make it across equilibrium to the other side. Keep this in mind for the next semester, where *exactly the same relationship*

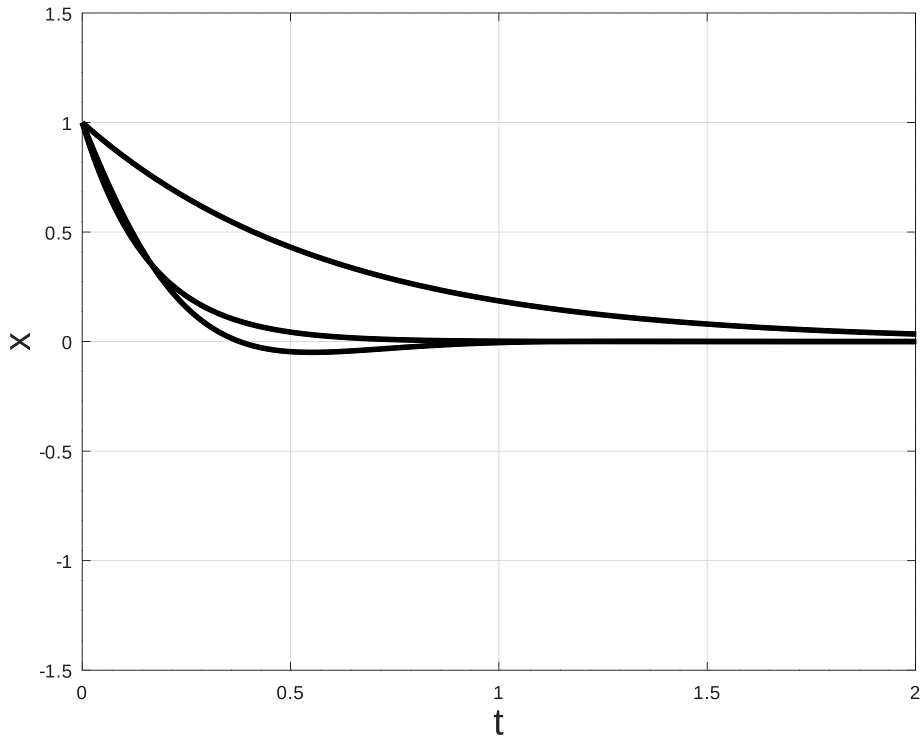


Figure 9.10: Three curves: Underdamped ($b/m = 2\pi$) barely oscillates. T' is now clearly longer than T_0. Critically damped ($b/m = 4\pi$) goes exponentially to zero in minimum time. Overdamped ($b/m = 8\pi$) goes to zero exponentiall, but much more slowly.

exists for LRC circuits, which exhibit damped simple harmonic oscillation that is *precisely the same* as that seen here for a linearly damped mass on a perfect spring.

Example 9.4.1: Car Shock Absorbers

This example isn't something one can compute, it is something you experience nearly every day, at least if you drive or ride around in a car.

A car's shock absorbers are there to reduce the "bumpiness" of a bumpy road. Shock absorbers are basically big powerful springs that carry your car suspended in equilibrium between the weight of the car and the spring force.

If your wheels bounce up over a ridge in the road, the shock absorber spring compresses, storing the energy from the "collision" briefly and then giving it back without the car *itself* reacting. However, if the spring is not damped, the subsequent motion of the car would be to bounce up and down for many tens or hundreds of meters down the road, with your control over the car seriously compromised. For that reason, shock absorbers are strongly damped with a suitable fluid (basically a thick oil).

If the oil is *too* thick, however, the shock absorbers become overdamped. The car takes so long to come back to equilibrium after a bump that compresses them that one rides with one's shocks constantly somewhat compressed. This reduces their effectiveness and one feels “every bump in the road”, which is *also* not great for safe control.

Ideally, then, your car’s shocks should be ***barely underdamped***. This will let the car bounce *through* equilibrium to where it is “almost” in equilibrium even faster than a perfectly critically damped shock and yet still rapidly damp to equilibrium right away, ready for the next shock.

So here’s how to test the shocks on your car. Push down (with all of your weight) on each of the corner fenders of your car, testing the shock on that corner. Release your weight suddenly so that it springs back up towards equilibrium.

If the car “bounces” *once* and then returns to equilibrium when you push down on a fender and suddenly release it, the shocks are good. If it bounces three or four time the shocks are too underdamped and dangerous as you could lose control after a big bump. If it doesn’t bounce up and back down at all at all and instead slowly oozes back up to level from below, it is overdamped and dangerous, as a succession of sharp bumps could leave your shocks still compressed and unable to absorb the impact of the last one and keep your tires still on the ground.

Damped oscillation is ubiquitous. Pendulums, once started, oscillate for only a while before coming to rest. Guitar strings, once plucked, damp down to quiet again quite rapidly. Charges in atoms can oscillate and give off *light* until the self-force exerted by their very radiation they emit damps the excitation. Cars need barely underdamped shock absorbers. Very tall buildings (“skyscrapers”) in a city usually have specially designed dampers in them as well to keep them from swaying too much in a strong wind. Houses are build with lots of damping forces in them to keep them quiet. Fully understanding damped (and eventually driven) oscillation is essential to many sciences as well as both mechanical and electrical engineering.

9.4.2: Energy Damping: Q-value

Suppose that one has a “weakly” damped oscillator. In the terminology above, that means that it is underdamped and, specifically, that $\omega' \approx \omega_0$ (because $1 \gg \frac{b^2}{4km}$). Every full oscillation of the system results in a loss of energy to the nonconservative work of the linear damping force. We can characterize this loss of energy as the ***dimensionless fraction of energy lost per cycle***:

$$\frac{\Delta E}{E} \tag{9.105}$$

We can easily compute this in the specific limit of the weakly damped oscillator. Let’s assume that we start the oscillator at any point where it has a maximum amplitude (and hence with zero phase) so that (for $\omega' \approx \omega_0$, recall):

$$x(t) = Ae^{-bt/2m}\cos(\omega_0 t) \tag{9.106}$$

Then at $t = 0$, there is no kinetic energy and:

$$E_0 = U_0 = \frac{1}{2}kA^2 \tag{9.107}$$

After one cycle (at time $t = T_0 = 2\pi/\omega_0$):

$$E_f = U_f = \frac{1}{2}kA^2 e^{-bT_0/m} \tag{9.108}$$

or:

$$\frac{\Delta E}{E_0} \frac{E_f - E_0}{E_0} = \frac{\frac{1}{2}kA^2\left(e^{-bT_0/m} - 1\right)}{\frac{1}{2}kA^2} = \left(e^{-bT_0/m} - 1\right) \tag{9.109}$$

This result is actually independent of any particular cycle or phase angle – it is the dimensionless fractional energy loss per cycle.

Weak damping specifically implies that $bT_0/m \ll 1$, so we can expand the exponential, cancel the leading 1, and keep the first surviving term:

$$\frac{\Delta E}{E_0} = ((1 - bT_0/m + ...) - 1) \approx \frac{-bT_0}{m} \tag{9.110}$$

This dimensionless relation looks pretty useful in and of itself, but (as we shall see) a slightly different form is more useful for describing damped *driven* harmonic oscillators. To obtain this form, we express $T_0 = 2\pi/\omega_0$ and rearrange, defining the result to be the ***Q-value*** of the damped oscillator:

$$Q = 2\pi\frac{E}{|\Delta E|} = \frac{m\omega_0}{b} \tag{9.111}$$

(Note that we lose the minus sign in the process – we just remember that this is related to the rate at which the oscillator ***loses*** energy.)

Note that the Q-value is basically $1/2\zeta$ as defined above, where the factor of 2 arises because the energy goes as position *squared*. It has the opposite meaning of ζ as well. ***Weak*** damping is ***large*** Q or *small* ζ. ***Strong*** damping is ***small*** Q or *large* ζ.

In English, the "Q" stands for "quality", and the Q factor is also called the *quality* factor. This now makes verbal sense – a "high quality oscillator" is one that oscillates a long time with slow damping and has large Q-for-quality compared to one that rapidly damps.

9.5: Damped, Driven Oscillation: Resonance

By and large, most of you who are reading this textbook directly experienced ***damped, driven oscillation*** long before you were five years old, in some cases as early as a few months old! This is the physics of ***the swing***, among other things. Babies love swings – one of our sons was colicky when he was very young and would sometimes only be able to get some peace (so we could get some too!) when he was tucked into a wind-up swing. Humans of all ages seem to like a rhythmic swaying motion; children play on swings, adults rock in rocking chairs.

Damped, driven, oscillation is also key in another nearly ubiquitous aspect of human life – the *clock*. Nature provides us with a few "natural" clocks, the most prominent one being the diurnal clock associated with the rotation of the Earth, read from observing the orientation of the sun, moon, and night sky and translating it into a time.

The human body itself contains a number of clocks including the most accessible one, the heartbeat. Although the historical evidence suggests that the size of the second is derived from systematic divisions of the day according to numerological rules in the extremely distant past, surely it is no accident that the smallest common unit of everyday time almost precisely matches the human heartbeat. Unfortunately, the "normal" human heartbeat varies by a factor of around three as one moves from resting to heavy exercise, a range that is further increased by the *ab*normal heartbeat of people with cardiac insufficiency, cardioelectric abnormalities, or taking various drugs. It isn't a very *precise* clock, in other words, although as it turns out it played a key role in the development of precise clocks, which in turn played a *crucial* role in the invention of physics.

Here, there is an interesting story. Galileo Galilei used his own heartbeat to time the oscillations of a large chandelier in the cathedral in Pisa around 1582 and discovered one of the key properties of the oscillations of a pendulum[223] (discussed above), *isochronism*: the fact that the period of a pendulum is independent of both the (small angle) amplitude of oscillation and the mass that is oscillating. This led Galileo and a physician friend to invent both the *metronome* (for musicians) and a simple pendulum device called the *pulsilogium* to use to time the pulse of patients! These were the world's first really accurate clocks, and variations of them eventually became the *pendulum clock*. Carefully engineered pendulum clocks that were compensated for thermal expansion of their rods, the temperature-dependent weather dependent buoyant force exerted by the air on their pendulum, friction and damping were the best clocks in the world and used as international time standards through the late 1920s and early 1930s, when they were superceded by *another*, still more accurate, damped driven oscillator – the quartz crystal oscillator.

Swings, springs, clocks and more – driven, resonant harmonic oscillation has been a part of everyday experience for at least two or three thousand years. Although it isn't obvious at first, *ordinary walking* at a comfortable pace is an example of damped, driven oscillation and resonance.

Military marching – the precise timing of the pace of soldiers in formation – was apparently invented by the Romans. Indeed, this invention gives us one of our most common measures of distance, the ***mile***. The word mile is derived from *milia passuum* – a thousand paces, where a "pace" is a complete cycle of two steps with a length slightly more than five feet – and Roman armies, by marching at a fixed "standard" pace, would consistently cover twenty miles in five summer hours, or by increasing the length of their pace slightly, twenty four miles in the same number of hours. Note well that the mile was originally a *decimal quantity* – a multiple of ten units! Alas, the "pace" did not become the unit of length in England – the "yard" was instead, defined (believe it or not) in terms of the width of a grain of barleycorn[224] . Yes folks, you heard it first here – there is an intimate connection between the *making of beer* (barley is one of the oldest cultivated grains and was used primarily for making beer and as an animal fodder dating back to *neolithic* times) and *the English units of length*.

The proper definition of a mile in the English system is thus the length of 190080 ***grains of barleycorn!***! That's almost exactly four cubic feet of barley, which is enough to make roughly 20 gallons of beer. Coincidence? I don't think so. And people wonder why the rest of the world

---

[223]Wikipedia: http://www.wikipedia.org/wiki/pendulum. I'd strongly recommend that students read through this article, as it is absolutely fascinating. At this point you should already understand that the development of physics required ***good clocks!*** It quite literally *could not have happened* without them, and good clocks, sufficiently accurate to measure e.g. the variations in the apparent gravitational field with height and position around the world, did not exist before the pendulum clock was conceived of and partially invented by Galileo in the late 1500s and invented in fact by Christiaan Huygens in 1656. Properties of the motion of the pendulum were key elements in Newton's invention of both the law of gravitation and his physics.

[224]Wikipedia: http://www.wikipedia.org/wiki/Yard. Yet another fascinating article – three barley grains to the inch, twelve inches to the foot, three feet to the yard, and $22 \times 220 = 4840$ square yards make an acre.

considers Americans and the British to be mad...

Roman soldiers also discovered another important aspect of resonance – it can *destroy human-engineered structures!* The “standard marching pace” of the Roman soldiers was 4000 paces per hour, just over two steps per second. This pace could easily *match* the natural frequency of oscillation of the *bridges* over which the soldiers marched, and driving a bridge oscillation, at resonance, with the weight of a hundred or so men was more than enough to ***destroy the bridge***. Since Roman times, then, although soldiers march with discipline whenever they are on the road, they *break cadence* and cross bridges with an irregular, random step lest they find themselves and the remnants of the bridge in the water, trying to swim in full armor.

This sort of resonance also affects the stability of buildings and bridges today – earthquakes can drive resonances of either one, the wind can drive resonances of either one. Building a sound bridge or tall building requires a careful consideration and damping of the natural frequencies of the structure. The ***Tacoma Bridge Disaster***[225] serves as a modern-times example of the consequences of failing to design for resonance. In more recent times part of the devastation caused by the Haiti Earthquake[226] was caused by the lack of earthquake-proofing – protection against earthquake-driven resonances – in the cheap construction methods used in buildings of all sorts.

From all of this, it seems like establishing at the very least a semi-quantitative understanding of resonance is in order, and as usual math, physics or engineering students will need to go the extra 190080 barleycorn grain lengths and work through the math properly. To manage this, we need to begin with a model.

9.5.1: Harmonic Driving Forces

Let’s start by thinking about what you all know from learning to swing on a swing. If you just sit on a swing, nothing happens. You have to “pump” the swing. Pumping the swing is accomplished by pulling the ropes and shifting your weight at the highest point of each oscillation so that the force exerted by the rope no longer passes through your center of mass and hence can *exert a torque in the current direction of rotation*. You know from experience that this torque must be applied *periodically* at a frequency that *matches* the natural frequency of the swing and *in phase with the motion* of the swing in order to increase the amplitude of the swinging oscillation. If you simply jerk backwards and forwards with the same motions as those used to pump “right” but at the wrong (non-resonant) frequency or randomly, you don’t ever build up much amplitude. If you apply the torques with the wrong *phase* you will not manage to get the same amplitude that you’d get pumping in phase with the motion.

That’s really it, qualitatively. Resonance consists of driving an oscillator at its natural frequency, in phase with the motion, to achieve the greatest possible oscillation amplitude (ultimately limited by things like practical physical constraints and damping).

[225] Wikipedia: http://www.wikipedia.org/wiki/Tacoma Narrows Bridge (1940). Again, a marvelous article that contains a short clip of the bridge oscillating in resonance to collapse. Note that the resonance in question was due more to the driving of several ***wave*** resonances rather than a simple harmonic oscillator resonance, but the principle is exactly the same.

[226] Wikipedia: http://www.wikipedia.org/wiki/2010 Haiti earthquake.

A pendulum, however, isn't a good system for us to use as a *quantitative* model. For one thing, it isn't really a harmonic oscillator – we automatically adjust our pumping to remain resonant as we swing closer and closer to the angle of $\pi/2$ where the swing chains become limp and we can no longer cheat some torque out of the combination of the pivot force and gravity, but the frequency itself starts to significantly change as the small angle approximation breaks down, and breaking it down is the *point* of swinging on a swing! Who wants to swing only through small angles!

The kind of force we exert in swinging a swing isn't too great, either. It is hardly smooth – we really only pump at all very close to the top of our swinging motion (on both sides) – in between we just coast. We'd prefer instead to assume a periodic driving force that is mathematically relatively easy to treat.

These two things taken together more or less uniquely determine the best model for us to use to understand resonance. This is an ***underdamped mass on a spring*** being driven by an external ***harmonic driving force***, all in one dimension:

$$F_x^{\rm ext}(t) = F_0 \cos(\omega t) \tag{9.112}$$

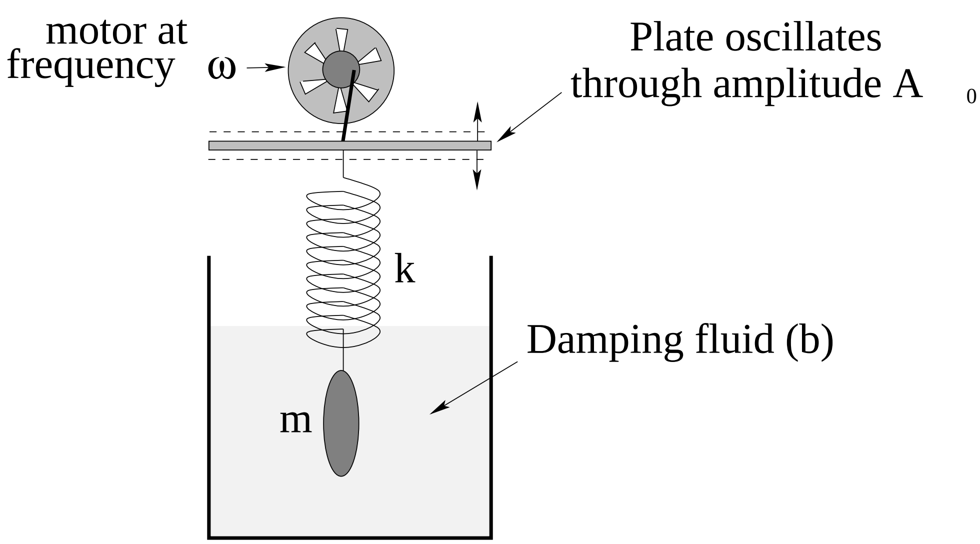


Figure 9.11: A small frequency-controlled motor drives the "fixed" end of a spring attached to a damped mass up and down through a (variable) amplitude A_0 at (independently adjustable) angular frequency ω, thereby exerting an additional harmonic driving force on the mass.

Although there are a number of ways one can exert such a force in the real world, one relatively simple one is to drive the "fixed" end of the spring harmonically through some amplitude e.g. $A(t) = -A_0 \cos(\omega t)$; this will modulate the total force exerted by the spring on the mass in just the right way:

$$\begin{aligned} F_x(t) &= -k(x + A(t)) - bv \\ &= -kx - bv + kA_0 \cos(\omega t) \\ &= -kx - bv + F_0 \cos(\omega t) \end{aligned} \tag{9.113}$$

where we've included the usual linear damping force $-bv$. The minus sign is chosen deliberately to make the driving force positive on the right in the *in*homogeneous ODE obtained below. The apparatus we might use to observe this under controlled circumstances in the lab is drawn in figure 9.11.

Newton's Second Law is then:

$$F_x(t) = -kx - bv + F_0 \cos(\omega t) = ma \tag{9.114}$$

If we write v and a as derivatives of $x(t)$, divide both sides by m, and rearrange, we can (as usual) turn this into a ***second order, linear, inhomogeneous ordinary differential equation*** (IODE) of motion in standard form and order:

$$\frac{d^2x}{dt^2} + \frac{b}{m}\frac{dx}{dt} + \frac{k}{m}x = \frac{F_0}{m}\cos(\omega t) \tag{9.115}$$

Note well that ω in the driving force is *not* necessarily equal to or related to either the angular frequency of the undamped oscillator $\omega_0 = \sqrt{k/m}$ or the shifted angular frequency of the damped oscillator $\omega' = \omega_0\sqrt{1 - b^2/4km}$ obtained above. Both ω and F_0 are "knobs" on motor driving the oscillator, that can be set to any value (within reason) to drive the damped oscillator at that frequency and with that magnitude of external force.

There are a number of ways to solve this differential equation of motion. A commonly shown one is graphical and fairly simple – making a "guess" as to the form of the steady-state solution (discussed below), transforming the resulting equation into a special kind of "vector" called "phasors"[227] to make a "phasor diagram" from which one can obtain the critical details of the solution using a mix of geometry and trigonometry.

It is, however, more elegant (and in the long run, more general and more correct) to proceed along the lines we have already explored for solving the undamped and damped undriven SHOE – to use complex exponentials combined with the Euler relation to transform the ODE into a relatively straightfoward algebra problem.

If you are not a math or physics major, or interested engineering student (if you plan to be an electrical engineer, you *should* be interested as this approach is ultimately the basis of a major fraction of electrical engineering as we'll see in the electricity and magnetism part of the course in the future) you most likely won't need to know how to solve the ODE and can safely skip over most of the next subsection to the place where it tells you to resume reading.

All students should come out of this section understanding what resonance *is*, how it works, and be able to semi-quantitatively sketch the power curve for various given values of Q. All students should know how Q varies with k, m, and b. All students should be able to answer conceptual questions involving resonance. Even life-science students, note well, should understand resonance well enough to be able to make sense of *magnetic resonance imaging*, and resonance is a major factor in the purely human activity of making music, so the effort spent here will not be wasted.

You can skip, skim, or study in detail as desired or required from here until the next horizontal line in the text.

---

[227] Sadly, these phasors can't be used to shoot aggressive space aliens they way they can in Star Trek...

9.5.2: Solution to Damped, Driven SHO

We start with the IODE with a *harmonic driving (inhomogeneous) term*:

$$\frac{d^2x}{dt^2} + \frac{b}{m}\frac{dx}{dt} + \frac{k}{m}x = \frac{F_0}{m}\cos(\omega t) \tag{9.116}$$

Finding the solution is *still* effectively integrating twice. Consequently, there ought to be two constants of integration in the final solution that can be used to match it to any set of given initial conditions or equivalent constraints.

We can easily identify where this freedom to match initial conditions come from. Suppose $x_{ss}(t)$ is a solution to the full inhomogeneous ODE – in math textbooks this solution is often called a *particular* solution, but in this context we will identify it as the "steady state" solution observed after a long time *independent* of the initial conditions, and have so labelled it with an "ss".

It is easy to verify (from the linearity of the ODE on the left) that we can always add a solution to the associated **homogeneous** ODE:

$$\frac{d^2x_{tr}}{dt^2} + \frac{b}{m}\frac{dx_{tr}}{dt} + \frac{k}{m}x_{tr} = 0 \tag{9.117}$$

to the original solution and we will still have a solution!

Fortunately, this isn't any burden to us; this is just the usual *damped simple harmonic oscillator equation*, so this part of the solution is *transient* and will eventually *decay exponentially to zero*. I have therefore labelled *it* accordingly with a "tr" for "transient". This means that a completely general solution can *always* be written as the sum of these two parts:

$$x(t) = x_{tr}(t) + x_{ss}(t) = Ae^{-bt/2m}\cos(\omega' t + \phi) + x_{ss}(t) \tag{9.118}$$

where now it is apparent that the $x_{ss}(t)$ is indeed "steady state" solution after the transient part has died away. A and ϕ are *as usual* the free constants of integration that can be used to ensure that the initial conditions are met, but we are left with the burden of finding $x_{ss}(t)$.

Now, we follow the same general methodology we used before, but instead of assuming a real exponential form and being forced into a detour through complex exponentials, we'll go ahead and make the entire IODE complex from the beginning, expecting to have to take the real part at the end as before. In particular, we will make the *driving force* complex in just the right way that the real part is the actual driving force in the original IODE, e.g.

$$\frac{F_0}{m}\cos(\omega t) = \Re e \frac{F_0}{m}e^{i\omega t}|$$

That is, the IODE becomes:

$$\frac{d^2x_{ss}}{dt^2} + \frac{b}{m}\frac{dx_{ss}}{dt} + \frac{k}{m}x_{ss} = \frac{F_0}{m}e^{i\omega t} \tag{9.119}$$

and assume a complex time dependent exponential form *with the same driving frequency* ω for the steady state solution $x_{ss}(t)$. We also make the amplitude of the solution a complex number as before:

$$x_{ss} = Xe^{i\delta}e^{i\omega t} \tag{9.120}$$

Note well that X and δ in this equation are real, and that they are *not* free parameters to be set from initial conditions. In fact, we have to solve for their unique values that satisfy the algebraic equations we obtain below! Also note once again that ω is the *given* frequency of the driving force, *not* related to ω' or ω_0 for the undriven oscillator!

Substituting and taking the derivatives, we get:

$$\left(-\omega^2 + i\frac{b}{m}\omega + \frac{k}{m}\right) X e^{i\delta} e^{i\omega t} = \frac{F_0}{m} e^{i\omega t} \tag{9.121}$$

We can now ***divide out the complex time dependence***, turning the IODE into a complex algebraic equation[228]:

$$\left(-\omega^2 + i\frac{b}{m}\omega + \frac{k}{m}\right) X e^{i\delta} = \frac{F_0}{m} \tag{9.122}$$

Now, as noted above, we have to solve for $Xe^{i\delta}$, as everything else in this equation is given. Noting that $k/m = \omega_0^2$ and dividing:

$$Xe^{i\delta} = \frac{F_0}{\left((\omega_0^2 - \omega^2) + i\frac{b}{m}\omega\right) m} \tag{9.123}$$

We have to turn the right hand side into a complex number, which means that we have to move the complexity from the denominator to the numerator. This is accomplished by rationalizing – we multiply the numerator and denominator both by the complex conjugate of the denominator. We then do a bit of rearranging to isolate and identify the amplitude and the phase:

$$\begin{aligned} Xe^{i\delta} &= \frac{F_0\left((\omega_0^2 - \omega^2) - i\frac{b}{m}\omega\right)}{\left((\omega_0^2 - \omega^2)^2 + \frac{b^2}{m^2}\omega^2\right) m} \\ &= \frac{F_0}{m\left((\omega_0^2 - \omega^2)^2 + \frac{b^2}{m^2}\omega^2\right)^{\frac{1}{2}}} \times \left\{ \frac{\left((\omega_0^2 - \omega^2) - i\frac{b}{m}\omega\right)}{\left((\omega_0^2 - \omega^2)^2 + \frac{b^2}{m^2}\omega^2\right)^{\frac{1}{2}}} \right\} \end{aligned} \tag{9.124}$$

If we let:

$$\cos\delta = \frac{\left((\omega_0^2 - \omega^2)\right)}{\left((\omega_0^2 - \omega^2)^2 + \frac{b^2}{m^2}\omega^2\right)^{\frac{1}{2}}} \tag{9.125}$$

$$-i\sin\delta = -i\frac{\left(\frac{b}{m}\omega\right)}{\left((\omega_0^2 - \omega^2)^2 + \frac{b^2}{m^2}\omega^2\right)^{\frac{1}{2}}} \tag{9.126}$$

and use the Euler relation $e^{-i\delta} = \cos\delta - i\sin\delta$, we can identify:

$$X = \frac{F_0}{m\sqrt{\left((\omega_0^2 - \omega^2)^2 + \frac{b^2}{m^2}\omega^2\right)}} \quad \text{and} \quad \delta = -\tan^{-1}\frac{\omega b}{(\omega_0^2 - \omega^2)m}$$

228 It might seem that we made a lot of guesses and assumptions in this conversion, but in an actual math course studying this you would *Fourier transform* the IODE and get to exactly this result without assumptions or dividing. This "math-ier" approach allows one in principle to handle more general, not necessarily harmonic, driving forces, but is wa-a-a-a-y beyond our scope here.

We can then assemble these parts:

$$x_{ss}(t) = Ae^{-bt/2m}\cos(\omega' t + \phi) + \mathrm{Re}\left(Xe^{i(\omega t - \delta)}\right) \tag{9.127}$$

(note where I put the negative phase, inside the exponential function argument $(\omega t - \delta)$) and finally ***take the real part*** of this to get the *real* solution we desire.

Resume Here, if you skipped the derivation...

The complete solution to the IODE for the damped, driven simple harmonic oscillator is therefore:

$$\boxed{x_{ss}(t) = Ae^{-bt/2m}\cos(\omega' t + \phi) + X\cos(\omega t - \delta)} \tag{9.128}$$

where:

$$\boxed{X = \frac{F_0}{m\sqrt{\left((\omega_0^2 - \omega^2)^2 + \frac{b^2}{m^2}\omega^2\right)}}} \quad \text{and} \quad \boxed{\delta = \tan^{-1}\frac{\omega b}{(\omega_0^2 - \omega^2)m}} \tag{9.129}$$

and the *constants of integration* A and ϕ, in the transient (exponentially decaying) part of the solution are available to make the solution satisfy the right/given initial (or other) conditions.

So much for the solution. Things you should know about it:

- It's made up of a transient part and steady state part.
- The transient part dies away exactly like a damped *undriven* harmonic oscillator.
- The steady state part is a simple harmonic oscillation ***at the same exact frequency as the harmonic driving force***, but...
- It's amplitude X and phase δ are not free parameters, but are rather entirely determined by F_0, k, m, and b (and the particular frequency ω of the driving force).
- The amplitude of steady state oscillation X peaks at, or very near, $\omega = \omega_0$, where one of the terms in the denominator vanishes.

As it turns out, we are not that interested, usually, in the amplitude per se. We are usually a lot more interested in the *power*, the rate that the driving force does *work* on the oscillator, work that is (in steady state) dissipated at precisely the same rate that it is added by the dissipative linear drag force only, as the spring itself is a conservative force. So let's examine that.

9.5.3: Power Delivered to the Driven Oscillator

Henceforth, we will assume that enough time has passed that the transient has died away. We will concentrate on $x_{ss}(t)$ only, in other words. The amplitude X of the steady state term is a ***maximum when $\omega \approx \omega_0$***. The power delivered to the system by the force F (next) is also a maximum. This is called ***resonance***.

To compute the time-dependent ***power*** delivered to the oscillator by the driving force we have to use (from Chapter/week 3):

$$P(t) = \vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{v}} = F(t) v_{ss}(t) \tag{9.130}$$

We take the time derivative of the steady state solution to get:

$$v_{ss}(t) = \frac{dx_{ss}(t)}{dt} = -\omega X \sin(\omega t - \delta) \tag{9.131}$$

This gives us:

$$P(t) = -F_0 \omega X \cos(\omega t) \sin(\omega t - \delta) \tag{9.132}$$

The quantity we really would like is the *average* power delivered to the oscillator. To evaluate this, we need to use the addition theorem for e.g. $\sin(A \pm B)$ to expand the $\sin(\omega t - \delta)$ However, chances are excellent that you (dear reader) *don't remember* this theorem/identity that you in principle learned back when you were taking trigonometry. For that reason, I'm going to insert a ***really useful trick*** that let's you ***derive*** the addition theorems for both cosine and sine in *two lines of algebra* using the Euler relation – effectively a one line derivation each!

Quick Derivation of Addition Theorems for $\sin$ and $\cos$

Line 1:

$$\cos(A+B) + i\sin(A+B) = e^{i(A+B)} = e^{iA}e^{iB} = (\cos A + i\sin A) \times (\cos B + i \sin B)$$

Then we distribute the product on the left and equate the real and imaginary parts.

Line 2:

$$\cos(A+B) + i\sin(A+B) = (\cos A \cos B - \sin A \sin B) + i(\cos A \sin B + \sin A \cos B)$$

From this we read off:

$$\cos(A+B) = (\cos A \cos B - \sin A \sin B)$$
$$\sin(A+B) = \cos A \sin B + \sin A \cos B$$

Note that cosine is an even function, and sine is odd, so (with $A = \omega t$ and $B = -\delta$):

$$\sin(A+B) = \cos(\omega t)\sin(-\delta) + \sin(\omega t)\cos(\delta) = \sin(\omega t)\cos(\delta) - \cos(\omega t)\sin(\delta) \tag{9.133}$$

Substituting this into the equation for instantaneous power we get:

$$P(t) = -F_0 \omega X \left\{ \cos(\omega t)\sin(\omega t)\cos(\delta) - cos^2(\omega t)\sin(\delta) \right\} \tag{9.134}$$

Now it is easy to average over time. On your own you should be able to show that averaging over a single period:

$$\frac{1}{T}\int_0^T \cos(\omega t)\sin(\omega t)dt = 0$$
$$\frac{1}{T}\int_0^T \cos^2(\omega t)dt = \frac{1}{2}$$

Averaging over *many* periods will always converge to these two values even if one doesn't integrate to/average over a strict integer multiple of periods.

Thus:

$$P_{\rm avg} = \frac{1}{2}\frac{F_0^2\omega\sin(\delta)}{m\,\sqrt{(\omega_0^2-\omega^2)^2+(b\omega/m)^2}} \tag{9.135}$$

and, substituting for $\sin\delta$:

$$\boxed{P_{\rm avg}(\omega) = \frac{F_0^2\omega^2 b}{2m^2\left\{(\omega_0^2-\omega^2)^2+(b\omega/m)^2\right\}}} \tag{9.136}$$

Note Well: I will ***not require you to know/memorize*** this formula *quantitatively* as it stands. (You should still go over this derivation in detail on your own!) However, you ***must*** be able to sketch what this function looks like semi-quantitatively for given (usually fairly low) values of Q! In order to do this efficiently, observe that:

$$P_{\rm avg}(0) = 0 \tag{9.137}$$
$$P_{\rm avg,max} = P_{\rm avg}\omega_0 = \frac{F_0^2}{2b} \tag{9.138}$$
$$\lim_{\omega\gg\omega_0} P_{\rm avg}(\omega) \approx \frac{F_0^2 b}{2m^2\omega^2} \to 0 \quad \text{like } \frac{1}{\omega^2} \tag{9.139}$$

If one sets:

$$P_{\rm avg}(\omega) = \frac{F_0^2\omega^2 b}{2m^2\left\{(\omega_0^2-\omega^2)^2+(b\omega/m)^2\right\}} = \frac{P_{\rm avg,max}}{2} = \frac{F_0^2}{4b} \tag{9.140}$$

then in the weak damping limit, two roots $\omega_\pm$ will emerge that are both very close to ω_0. In an advanced homework problem, you can show that the *difference* between these two roots – called the ***full (resonance curve) width at half-maximum power*** is:

$$\Delta\omega = \omega_+ - \omega_- = \frac{b}{m} \tag{9.141}$$

If you recall the ***quality factor*** of this damped oscillator Q and substitute in this result you get:

$$\frac{m\omega_0}{b} = \frac{\sqrt{km}}{b} = \boxed{Q = \frac{\omega_0}{\Delta\omega}} \tag{9.142}$$

We find that Q is not *only* (2π times) the inverse of the fractional energy loss per cycle, it is also the ratio of the full width at half-maximum to the resonant frequency ω_0! The three limits above, plus this result, make it quite simple to sketch a resonance curve that has these key features, given Q, ω_0, and $P_{\rm max}$.

Here is a set of steps for drawing a resonance curve, given Q, onto a pair of axes representing average power as a function of ω.

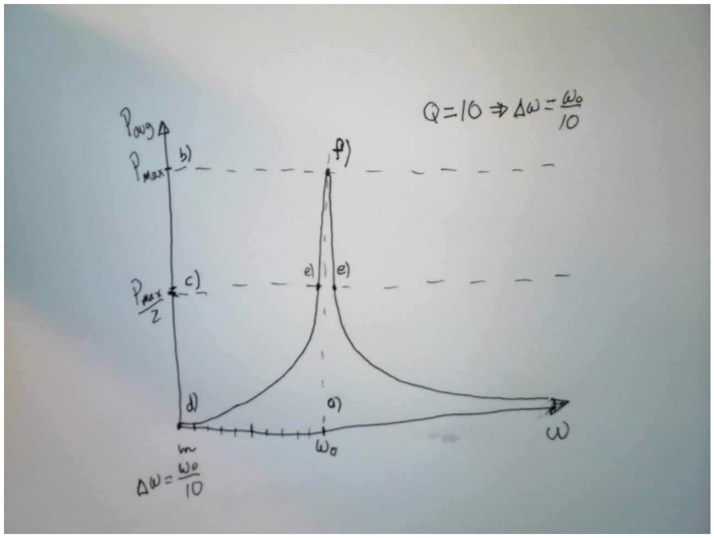


Figure 9.12: A hand-drawn sketch of a resonance curve for $Q = 10$. Note the labelled steps/points used to make the figure.

a) Draw a very faint dashed line vertically up from ω_0 on the ω axis.

b) Mark or locate $P_{\rm max}$ on the P axis and draw a very faint dashed line horizontally over.

c) Mark/locate $P_{\rm max}/2$ on the P axis and draw a very faint dashed line horizontally over.

d) *Estimate* $Q = \dfrac{\omega_0}{\Delta\omega} \;\Rightarrow\; \Delta\omega = \dfrac{\omega_0}{Q}$

e) Make two faint dots on the $P_{\rm max}/2$ line a distance $\Delta\omega/2$ to either side of the central line at ω_0.

f) Draw a "resonance-shaped curve" that starts at 0 at $\omega = 0$, rises *through* the first dot up to $P_{\rm max}$ at ω_0, falls *through* the second dot to fall off towards 0 like (eventually) $1/\omega^2$.

These steps are illustrated in figure 9.12. We can compare this hand-made sketch to actual computer-generated graphs of three resonance curves for $Q = 3, 10, 20$, plotted in figure 9.13. **Note Well:** Qualitatively, **Big *Q* is sharp and high**, while **small *Q* is short and fat**. **Big *Q*** means you are delivering a *lot of power to the oscillator at its natural frequency* and the amplitude of oscillation and everything tied to it is maximum! Big Q is also ***weak damping***, and really, our trick for drawing the resonance curve and identification of $\Delta\omega$ (made with some approximations in the advanced homework derivation) will be increasingly inaccurate for Q smaller than (say) 5 or so.

The point of understanding Q pretty thoroughly is that oscillators with low Q quickly damp and don't build up much amplitude even from a perfectly resonant $\omega = \omega_0$ driving force. This is "good" when you are building bridges and skyscrapers. High Q means you get a *large* amplitude, eventually, even from a *small* but perfectly resonant driving force. This is "good" for jackhammers, musical instruments, understanding a good walking pace, and many other things.

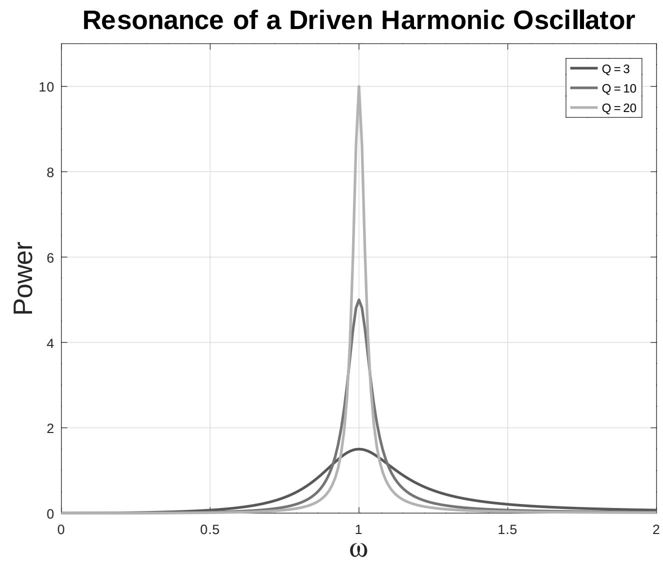


Figure 9.13: This plots the ***average power*** $P(\omega)$ for three resonances. In these figures, $F_0 = 1$, $k = 1$, $m = 1$ (so $\omega_0 = 1$) and hence b is the only variable. Since $Q = \omega_0/\Delta = m\omega_0/b$ we plot $Q = 3, 10, 20$ by selecting $b = 0.333, 0.1, 0.05$. Thus $P_{\rm max} = F_0^2/2b = 3/2, 5, 10$ respectively (all in suitable units for the quantities involved). Note that the full width of e.g. the $Q = 20$ curve (with the sharpest/highest peak) is $\sim 1/20$ at $P = P_{\rm max}/2 = 5$. P is the average power *added to* the oscillator by the driving force, which in turn in steady state motion must equal the average power *lost to* to dissipative drag forces.

9.6: Adding Springs in Series and in Parallel

At this point, you should have a pretty good understanding of how an ideal spring will make a mass oscillate (more or less) harmonically, with or without weak linear damping. You should also have the glimmerings of an idea why they are so important – almost *any* system in the near neighborhood of a stable equilibrium point will a very good chance of experiencing linear restoring forces in that neighborhood and hence is likely to make the mass oscillate harmonically around the equilibrium point.

However, we still have a mystery or two to address. One of the most important ones is: ***Why do springs behave like springs?*** Isn't there a bit of a chicken and egg problem here? Granted that linear springs make things oscillate harmonically, why are springs – things made up (we profoundly believe) of many, many atoms bound together by complicated interatomic forces – linear in the first place?

To understand that, as well as to address various issues with engineering devices that might contain more than one spring or use multiple springs at the same time to accomplish some design goal, we first need to learn how to *add up* springs in various arrangements and see how multiple springs in those arrangements will (perhaps surprisingly) behave like a single spring. Along the way, we will learn the spring-equivalents of "a chain is only as strong as its weakest link" and "E pluribus unum" – things that you might already understand just using common sense.

The two arrangements we will concentrate on are springs in ***series*** – one spring connected to another in a linear chain, where the springs are allowed to all have different spring constants – and in ***parallel*** where the springs are all side by side.

Our goal will be to determine algebraically what the spring constant of a *single* spring would have to be if we were to replace the entire series or parallel arrangement with that spring and observe the exact same restoring force for the exact same total stretch (or compression).

The algebra for this isn't terribly difficult, but for series springs in particular, it may not be very *intuitive* – you may need to practice to master the "just right" first steps to keep the algebra simple. It is important to master this algebra because as it turns out, almost *identical* reasoning and algebra is used to determine the total resistance of series and parallel arrangements of resistors or capacitors in electronic circuits covered in the next semester of this course. If you get it right now, and remember it then, it will make understanding these rules of electronics very simple when you get to them.

In the meantime, once we understand these rules for adding ideal springs, well be in a perfect position to understand why ***nearly all materials behave like springs*** when they are stretched, compressed, squeezed down in volume, or deformed. We will come to realize that even the good old normal force that we idealized way back at the beginning of the course is really just an example of this same principle – when we press our finger down on something, it *very slightly deforms* it, compressing tiny "intermolecular" springs that are then what exerts the observed normal force. When you stand on a ladder, the ladder gets *very slightly shorter* by an amount proportional to your weight and the length of the ladder, all due to these rules.

Let's get started.

9.6.1: Adding Springs in Series

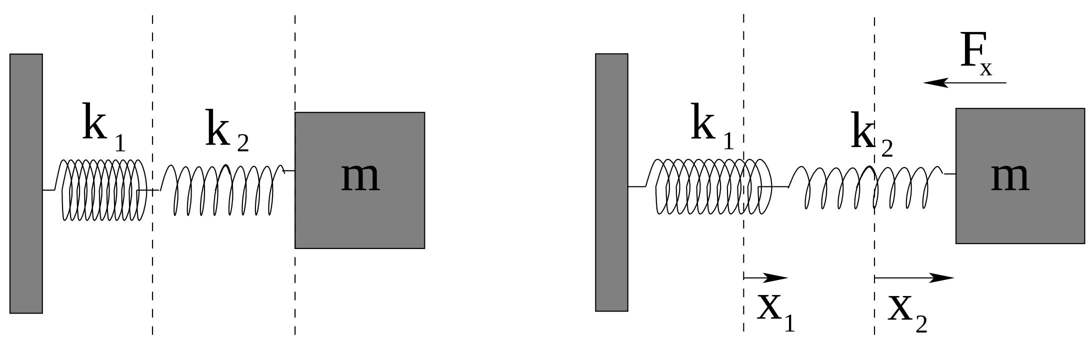


Figure 9.14: Two springs with spring constants k_1 and k_2 in *series*. The left-hand picture is before either spring is stretched, so that the mass is at equilibrium (no force due to springs). The right-hand picture shows the mass pulled out to the right so that the first spring has stretched a distance x_1 and the second (weaker) spring has stretched x_2.

The figure above illustrates what happens when only *two* springs are connected in series. When the attached mass is pulled out to the right and then held there, it stretches *both* of the two springs. Here's an important argument.

The total force acting on the mass due to the two springs is F_x, pulling back to its equilibrium position at the end of the two unstretched springs. From N3, that mass is pulling spring 2 to the right with an equal and opposite force. But spring 2 itself is in equilibrium! So spring 1 ***must*** be pulling it to the ***left*** with force F_x! This in turn means spring 2 is pulling spring 1 to the right with force F_x (N3 again) and spring 1 isn't moving, so spring 1 must be pulled to the left by the wall with F_x, and the wall is pulled to the right by F_x. Newton's Third Law is *useful!*

Clever students will recognize this is being very similar to our original arguments for why the tension in a stationary *string* is the same throughout the string, pulling any little differential chunk equally to the right and to the left and ultimately just transferring force from one end to the other. We've simply made a "string" out of our two springs!

This gives us a simple insight. We now know (from Hooke's Law and this reasoning) that the force exerted by either spring on its ends is:

$$F_x = -k_1 x_1 = -k_2 x_2$$

Now comes the tricky part. We'd like to find a *single* spring (with spring constant $k_{\rm eff}$) that produces this *same* force with the ***same total stretch***. That is, we'd like to find:

$$F_x = -k_{\rm eff} x$$

where

$$x_1 + x_2 = x$$

as shown in figure 9.15.

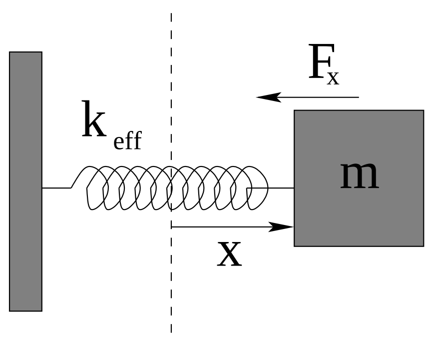


Figure 9.15: A single spring equivalent to the two in series (or the two in parallel, next).

This isn't all *that* tricky, but the first time you try this on your own, trust me, you are likely to try to ***incorrectly*** add $k_1 x_1$ to $k_2 x_2$ and get yourself into all kinds of algebraic difficulty. I certainly did, the first time I tried it (lo, those many years ago, and an embarrassing number of times afterwards). So don't feel bad if this is you; just learn the trick! *Series* springs add based on the *interesting* fact that ***the length every spring stretches adds up to a total length stretched***, while *the force is the same everywhere throughout the line of springs* and hence is a bit boring and not at all anything you can directly add up.

Boring or not, it is *useful*, so let's use it! We use it to express x_1, x_2, and x all in terms of the k's and the single F_x that is the *same* for both pictures and all three equations:

$$x_1 = F_x/k_1 \qquad x_2 = F_x/k_2 \qquad x = F_x/k_{\rm eff}$$

Now the remaining algebra is *simple!* We just substitute these three equations into the equation that represents the *interesting* fact regarding the sum of the stretches and cancel the common factor of F_x:

$$x_1 + x_2 = x \quad \Rightarrow \quad \frac{F_x}{k_1} + \frac{F_x}{k_2} = \frac{F_x}{k_{\rm eff}} \quad \Rightarrow \quad \frac{\cancel{F_x}}{k_1} + \frac{\cancel{F_x}}{k_2} = \frac{\cancel{F_x}}{k_{\rm eff}} \quad \Rightarrow \quad \boxed{\frac{1}{k_{\rm eff}} = \frac{1}{k_1} + \frac{1}{k_2}(+...)}$$

I took the liberty of adding a $(+...)$ to this last equation because it is hopefully pretty obvious that if I'd done *three*, or *four*, or N springs in series, they would all have the same force F_x but the sum of the lengths each spring stretched would now be $x_1 + x_2 + ...x_N = x$ (for any value of N, the total number of springs). This lets us write the sum rule for series in a nice, compact form:

$$\boxed{\frac{1}{k_{\rm eff}} = \sum_{i=1}^{N} \frac{1}{k_i}}$$

or in English, "the reciprocal of the total effective spring constant is the sum of the reciprocals of the spring constants in series".

Here's a single example of adding up three springs in series. The key is that to find $k_{\rm eff}$, you have to *put the sum over a common denominator and add them, and* ***only then*** *invert the result!* Watch out for this!

Example 9.6.1: Three Springs in Series

Suppose $k_1 = 1$, $k_2 = 2$, and $k_3 = 3$ (all in N/m). Then:

$$\frac{1}{k_{\rm eff}} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} = \frac{2 \times 3}{1 \times 2 \times 3} + \frac{1 \times 3}{1 \times 2 \times 3} + \frac{1 \times 2}{1 \times 2 \times 3} = \frac{6+3+2}{6} = \frac{11}{6}$$

or:

$$k_{\rm eff} = \frac{6}{11}\ \mathrm{N/m}$$

Note Well! A common mistake here would be to do:

$$\frac{1}{k_{\rm eff}} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} \Longrightarrow k_{\rm eff} = 1 + 2 + 3 = 6\ \mathrm{N/m}$$

This is very, very wrong! The reciprocal of a sum is not the same thing as the sum of the reciprocals! This is the *whole point* of putting things over a common denominator before adding them!

9.6.2: Adding Springs in Parallel

In parallel, the forces of the two springs just add. So:

$$F_x = -k_1 x - k_2 x = -(k_1 + k_2)x = -k_{\rm eff} x \Longrightarrow \boxed{k_{\rm eff} = k_1 + k_2(+...)}$$

and we can write this more generally as:

$$\boxed{k_{\rm eff} = \sum_{i=1}^{N} k_i}$$

for any number of springs in parallel.

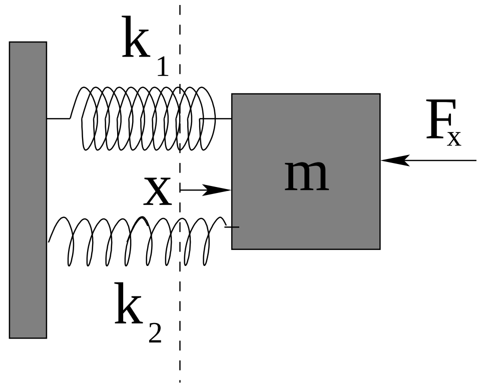


Figure 9.16: Two springs with spring constants k_1 and k_2 in *parallel*. This is much easier as it is obvious that *the forces of the springs just add for the common stretch* x*!*

Example 9.6.2: Adding Springs in Parallel

Suppose we are adding the same three springs – $k_1 = 1$, $k_2 = 2$, and $k_3 = 3$ (all in N/m) – but now we'e adding them in parallel. Now we really do get:

$$k_{\text{eff}} = 1 + 2 + 3 = 6 \text{ N/m}$$

Note that this is exactly what you would get if you ***misadded*** the three springs in series! How, you might wonder, can you keep this straight? On an exam, under stress, you have to recognize which is which, and you might need to reason conceptually about this in simple scaling problems as usual. This motivates the following short summary.

9.6.3: Rules of Thumb

If you look at the results above, when adding springs in series, the effective spring constant is always *smaller than the smallest spring constant in the series!* We have an adage for this, already quoted above: "a chain is only as strong as its weakest link". If one spring happens to be much weaker than all of the rest, most of the stretch will happen for that one weak spring, and it is likely to be the one that breaks if overstressed! However, even the stronger springs will stretch *a little*, so the total is *weaker* than this weakest spring.

A second rule for series is that if you have N *identical* springs in series:

$$\frac{1}{k_{\text{eff}}} = \frac{1}{k} + \frac{1}{k} + ... = \frac{N}{k}$$

then:

$$\boxed{k_{\text{eff}} = \frac{k}{N}}$$

This rule is a key component of the following section, so make sure that you understand it! N identical springs in series have a spring constant that is $1/N$th of the constant of any of the springs! A long line of springs gets *weaker* the longer it gets, at a rate *more or less proportional ot the total unstretched length!*

For parallel springs, it is obvious that the effective spring constant is ***larger than the largest*** spring constant in the parallel set. If one of the springs is super strong compared to the others (a spring used in the suspension of a car versus the springs you find on screen doors, for example) you have to stretch that car spring *and* you still have to exert enough additional force to stretch all of the weaker springs as well. This is just common sense. Adages include *E pluribus unum* – from many, one – used to suggest that together we are stronger than any of us is alone. However, the wisdom in this goes all the way back to elementary school, where everybody tried to pick the biggest, strongest kid to be on *their* team in a tug of war, where everybody's strength on a team gets *added*.

In a way, this is one of the motivations for *team based learning*, which is the way I always teach this course. A team of students working together has the double advantage of leveraging the *varied* strengths of its members *and* allowing them to share and pool that knowledge so that every member emerges from problem solving practice stronger than they went in!

Again we have a rule for adding M *identical* springs:

$$\boxed{k_{\text{eff}} = k + k + ... = M \times k}$$

You just multiply the spring constant by the number of springs.

These two rules, combined, are the entire basis of the next section. Make sure you understand them before proceeding.

9.7: Elastic Properties of Materials

It is now time to close a very important bootstrapping loop in your understanding of physics. From the beginning, we have used a number of force rules like "the normal force", and "Hooke's Law" because they were simple rules that we could directly observe and use to help us both understand ubiquitous phenomena and learn to use Newton's Laws to quantitatively describe many of them.

We had to do this *first*, because until you understand force, work, energy, equations of motion and conservation principles – basic mechanics – you cannot *start* to understand the microscopic basis for the macroscopic "rules" that govern both everyday Newtonian physics and things like thermodynamics and chemistry and biology (all of which have rules at the macroscopic scale that follow from physics at the microscopic scale).

We've done a bit of this along the way – thought about microscopic causes of friction and drag forces, derived the ways in which a macroscopic object can be thought of as a pointlike microscopic object located at its center of mass (and ways it cannot, e.g. when describing its rotation). We've even thought a *bit* about things like compressibility of fluids, but we haven't really thought about this *enough*.

Let's fix that.

To do so, we need a *microscopic model for a solid*, in particular for the molecular bonds within a solid.

9.7.1: Simple Models for Molecular Bonds

Consider, then a very crude microscopic model for a solid. We know that this solid is made up of *many* elementary particles, and that those elementary particles interact to form nucleons, which bind together to form nuclei, which bind elementary electrons to form atoms and that the atoms in turn are bound together by short range (nearest neighbor) interatomic forces – "chemical bond" if you like – to actually form the solid.

From both the text and some of the homework problems you should have learned that a "generic" potential energy associated with the interaction of a pair of atoms with a chemical bond between them is given by a ***short range repulsion*** followed by a ***long range attraction***. We saw a potential energy form *like* this as the "effective potential energy" in gravitation (the form that contained an angular momentum barrier with L^2 in it), but the physical origin of the terms is very different.

The repulsion in the molecular potential energy comes from first the ***Pauli exclusion principle***[229] in quantum mechanics (that makes the interpenetration of the electron clouds surrounding atoms energetically "expensive" as the underlying quantum states rearrange to satisfy it) plus the penetration of a ***screened Coulomb interaction***. The former is truly beyond the scope of this course – it is quantum magic associated with electrons[230] as fermions[231] , where I'm inserting wikipedia links to lots of these terms so that *interested* students can use them as the starting points of wikipedia romps, as a lot of this is all *absolutely fascinating* and is one of the reasons physicists love physics, it is all just so very amazing.

Next semester you *will* learn about ***Coulomb's Law***[232] , that describes the forces between two charged particles, and (using Gauss's Law[233]) you will be able to understand how the electron cloud normally "screens" the nuclei as eventually the rearrangement brings increasingly "bare" nuclei close enough so that the atoms have a *very* strong net repulsion.

One thing that we won't cover then, however, is how this simple/naive model, which leads to two atoms *not interacting at all* as soon as they are not "touching" (electron clouds interpenetrating) is replaced by one where two neutral atoms have a residual *long range interaction* due to dipole-induced dipole forces, leading to what is called a ***London dispersion force***[234] . This force has the generic form of an *attractive* $-C/r^6$ for a rather complicated C that parameterizes various details of the interatomic interaction.

Physicists and quantum chemists or engineers often idealize the exact/quantum theory with an approximate (semi)classical potential energy function that models these important generic features. For example, two very common models (one of which we already briefly explored in

---

[229] Wikipedia: http://www.wikipedia.org/wiki/Pauli Exclusion Principle.
[230] Wikipedia: http://www.wikipedia.org/wiki/Electron.
[231] Wikipedia: http://www.wikipedia.org/wiki/Fermion.
[232] Wikipedia: http://www.wikipedia.org/wiki/Coulomb's Law.
[233] Wikipedia: http://www.wikipedia.org/wiki/Gauss' Law.
[234] Wikipedia: http://www.wikipedia.org/wiki/London dispersion force. This force is due to ***Fritz London*** who was a Duke physicist of great reknown, although he derived this force from second-order perturbation theory long before he fled the rise of the Nazi party in Germany and eventually moved to the United States and took a position at Duke. London is honored with an special invited "London Lecture" at Duke every year. Just an interesting True Fact for my Duke students.

week 4, homework problem 4) is the ***Lennard-Jones potential***[235] (energy):

$$U_{LJ}(r) = U_{\min}\left\{2\left(\frac{r_b}{r}\right)^6 - \left(\frac{r_b}{r}\right)^{12}\right\} \tag{9.143}$$

In this function, $U_{\min}$ is the ***minimum*** of the potential energy curve (evaluated with the usual convention that if $r \to \infty$ then $U_{LJ}(\infty) = 0$), r_b is the radius where the minimum occurs (and hence is the equilibrium ***bond length***), and r is along the bond axis. This function is portrayed as the solid line in figure 9.17.

An alternative is the ***Morse potential***[236] (energy):

$$U_M(r) = U_{\min}\left(1 - e^{-a(r-r_b)}\right)^2 \tag{9.144}$$

In this expression $U_{\min}$ and r_b have the same meaning, but they are joined by a, a parameter than sets the *width* of the well. The only problem with this form of the potential is that it is difficult to compare it to the Lennard-Jones potential energy above because the LJ potential is zero at infinity and the Morse potential energy is $U_{\min}$ at in infinity. Of course, potential energy is *always* only defined within an additive constant, so I will actually subtract a constant $U_{\min}$ and display:

$$U_M(r) = -U_{\min} + U_{\min}\left(1 - e^{-a(r-r_b)}\right)^2) \tag{9.145}$$

as the dashed line in figure 9.17 below. This potential energy now "conventionally" vanishes at ∞.

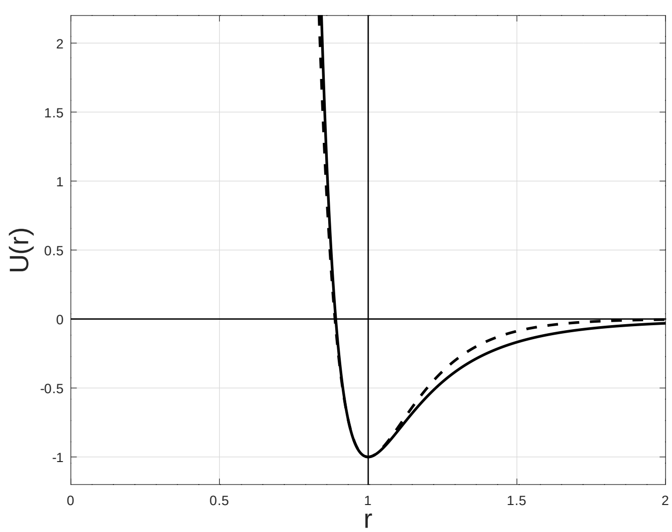


Figure 9.17: Two "generic" classical potential energy functions associated with atomic bonds on a common scale $U_{\min} = -1$, $r_b = 1.0$, and $a = 6.2$, the latter a value that makes the two potential have roughly the same ***force constant*** for small displacements.The solid line is the ***Lennard-Jones*** potential energy $U_L(r)$ and the dashed line is the ***Morse*** potential $U_M(r)$. Note that the two are very closely matched for short range repulsion, but the Morse potential dies off faster than the $1/r^6$ London form expected at longer range.

[235]Wikipedia: http://www.wikipedia.org/wiki/Lennard-Jones potential.
[236]Wikipedia: http://www.wikipedia.org/wiki/Morse potential.

I should emphasize that neither of these potential energies is in any sense a law of nature. They are effective potential energies, idealized and parameterized model potential energies that have close to the right shape and that can be used to study and understand molecular bonding in a semi-quantitative way – good enough to be compared to experimental results in an understandable if approximate way, but hardly exact.

The exact (relevant) laws of nature are those of electromagnetism and quantum mechanics, where the many electron problem must be solved, which is very difficult. The problem rapidly becomes too complex to really be solvable/computable as the number of electrons grows and more atoms are involved. So even in quantum mechanics people not infrequently work with Lennard-Jones or Morse potential energies (or any of a number of other related forms with more or less virtue for any particular problem) and sacrifice precision in the result for computability.

In the case we are interested in, however, even these relatively *simple* effective potential energies are too complex. We therefore take advantage of something I have written about extensively up above. Nearly *all* potential energy functions that have a true minimum (so that there is a force equilibrium there, recalling that the force is the negative slope (gradient) of the potential energy function) vary ***quadratically*** for small displacements from equilibrium. A quadratic potential energy corresponds to a harmonic oscillator and a linear restoring force. Therefore all solids made up of atoms bound by interactions like the effective molecular potential energies above will inherit certain properties from the tiny "springs" of the bonds between them!

9.7.2: The "Spring Constant" of a Molecular Bond

In both of the cases above, we can derive a quantity that acts as a sort of ***effective "spring constant"*** of the interatomic bonds that hold atoms in their place relative to a neighboring atom. The easiest way to do so is to do a ***Taylor Series Expansion***[237] of the potential energy function around r_b. This is:

$$U(x_0 + \Delta x) = U(x_0) + \left(\frac{dU}{dx}\right)\bigg|_{x=x_0} \Delta x + \frac{1}{2!}\left(\frac{d^2U}{dx^2}\right)\bigg|_{x=x_0} \Delta x^2 + ... \tag{9.146}$$

At a *stable equilibrium* point x_0 the force *vanishes*:

$$F_x(x_0) = \left(\frac{dU}{dx}\right)\bigg|_{x=x_0} = 0 \tag{9.147}$$

so that:

$$U(x_0 + \Delta x) = U(x_0) + \frac{1}{2!}\left(\frac{d^2U}{dx^2}\right)\bigg|_{x=x_0} \Delta x^2 + ... \tag{9.148}$$

We can now identify this with the ***form*** of a potential energy equation for a mass on a spring

[237] Wikipedia: http://www.wikipedia.org/wiki/Taylor series expansion.

in a coordinate system where the equilibrium position of the spring is at x_0[238]:

$$U(x_0 + \Delta x) = U(x_0) + \frac{1}{2!}\left(\frac{d^2U}{dx^2}\right)\bigg|_{x=x_0} \Delta x^2 + ... = U_0 + \frac{1}{2}k_{\rm eff}\Delta x^2 \quad (9.149)$$

where

$$k_{\rm eff} = \left(\frac{d^2U}{dx^2}\right)\bigg|_{x=x_0} \quad (9.150)$$

What this means is that ***any mass at a stable equilibrium point will usually behave like a mass on a spring for motion "close to" the equilibrium point!*** If pulled a short distance away and released, it will oscillate nearly harmonically around the equilibrium. The terms "usually" and "nearly" can be made more precise by considering the neglected third and higher order derivatives in the Taylor series – as long as they are negligible compared to the second order derivative with its quadratic Δx^2 dependence, the approximation will be a good one.

We have already seen this to be true and used it for the simple and physical pendulum problem, where we used a Taylor series expansion for the force or torque and for the energy and kept only the leading order term – the small angle approximation for $\sin(\theta)$ and $\cos(\theta)$. You will see it in homework problems later this semester and next semester as well (at least if you continue using this textbook) where you will from time to time use the binomial expansion (the Taylor series for a particular form of polynomial) to transform a force or potential energy associated with a particle near equilibrium into the form that reveals the simple harmonic oscillator within, so to speak. Any time you see a *linear* restoring force or torque, you can expect harmonic oscillation!

At *this* point, however, we wish to put this idea to a different use – to help us bridge the gap between microscopic forces that hold a "rigid" object together and that object's response to forces applied to it. After all, we *know* that there is no such thing as a truly rigid object. Steel is pretty hard, but with enough force we can bend "solid" steel, we can stretch or compress it, we can turn it into a spring, we can fracture it. Bone is also pretty hard and *it does exactly the same thing:* bend, stretch, compress, fracture. The physics of bending, stretching, compressing, and fracturing a solid object is associated with the application of a quantity called *stress* (with units of pressure, note well, but not *precisely* a pressure) to the object. Let's see how.

9.7.3: A Microscopic Picture of a Solid

Let us consider a solid piece of some simple material. If we look at pure elemental solids – for example metals – we often find that when they form a solid the atoms arrange themselves into a *regular lattice* that is "close packed" – arranged so that the atoms more or less touch each other with a minimum of wasted volume. There are a number of kinds of lattices that appear (determined by the subtleties of the quantum mechanical interactions between the atoms and hence beyond the scope of this course) but in many cases the lattice is a variant of the cubic lattice where there are atoms on the corners of a regular cartesian grid in three dimensions (and sometimes additional atoms in the center of the cubes or on the faces of the cubes).

[238]Remember, any potential energy function is defined only to within an additive constant, so the constant term U_0 simply sets the scale for the potential energy without affecting the actual force derived from the potential energy. The force is what makes things happen – it is, in a sense, all that matters.

Not all materials are so regular. Materials can be made out of a mixture of atoms, out of molecules made of a mixture of atoms, out of a mixture of molecules, or even out of living cells made out of a mixture of molecules. The resulting materials can be ordered, *structured* (not exactly the same thing as ordered, especially in the case of solids formed by life processes such as bone or coral), or disordered (amorphous).

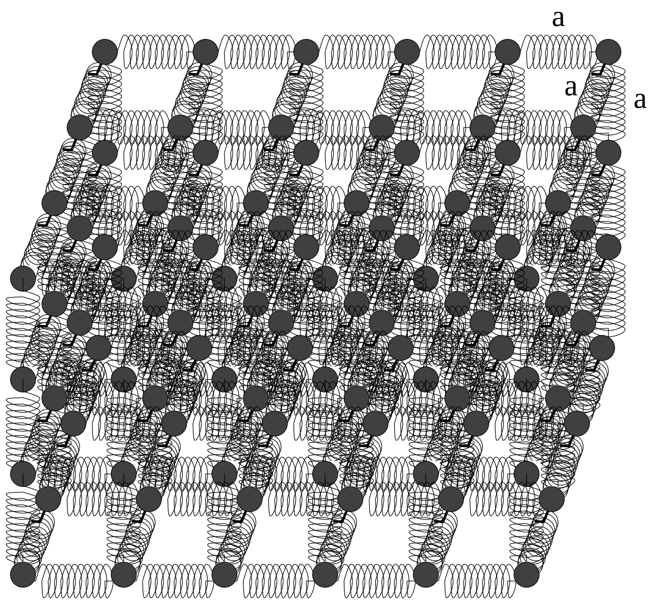


Figure 9.18: An idealized simple cubic lattice of atoms separated by the "springs" of interatomic forces that hold them in equilibrium positions. The equilibrium separation of the atoms as a.

As usual, we will deal with all of this complexity by ignoring most of it for now and considering an "ideal" case where a single kind of atoms lined up in a regular ***simple cubic lattice*** is sufficient to help us understand properties that will hold, with different values of course, even for amorphous or structured solids. This is illustrated in figure 9.18, which is basically a mental cartoon model for a generic solid – lots of atoms in a regular cubic lattice with a cube side a, where the interactomic forces that hold each atom in position is represented by a ***spring***, a concept that is valid as long as we don't compress or stretch these interatomic "bonds" by too much.

We need to quantify the numbers of atoms and bonds in a way that helps us understand how stretching or compressing forces are distributed among all of the bonds. Suppose we have N_x atoms in the x-direction, so that the length of the solid is $L = N_x a$. Suppose also that we have N_y and N_z atoms in the y- and z-directions respectively, so that the cross-sectional area $A = N_y N_z a^2$ (at least approximately, as we are not treating atoms on the boundary particularly accurately as they contribute to the problem at "lower order", meaning we can ignore them for large systems).

Now let us imagine applying a force (magnitude F) *uniformly* to all of the atoms on the left and right ends that *stretches* all of these bonds by a small amount Δx, presumed to be "small" in precisely the sense that leaves the interatomic bonds still behaving like springs. The force F has to be distributed equally among all of the atoms on the end areas on both sides, so that the force applied to each *chain* of atoms in the x-direction end to end is:

$$F_{\text{chain}} = \frac{F}{N_y N_z} = \frac{Fa^2}{A} \tag{9.151}$$

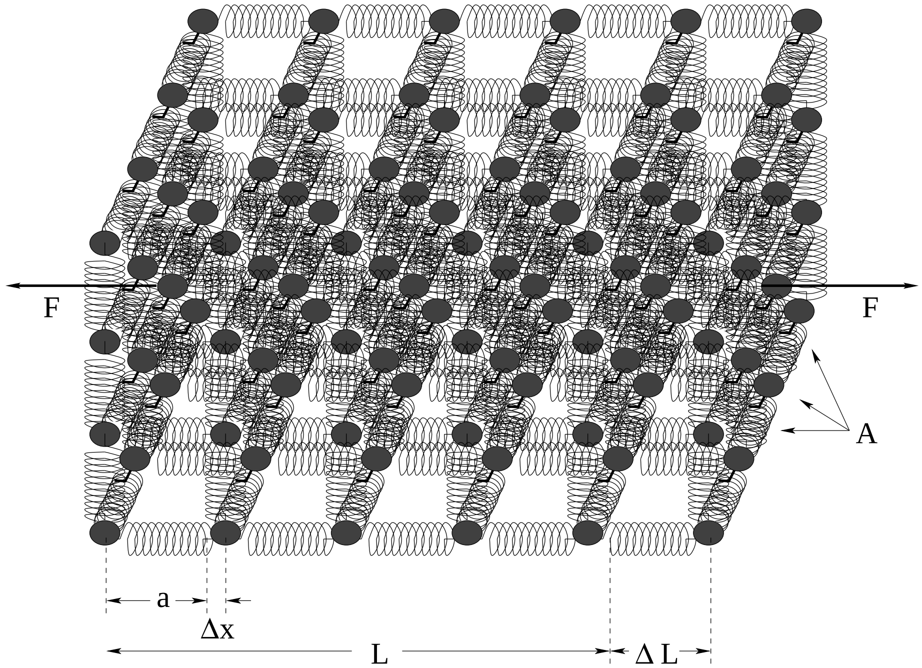


Figure 9.19: The same lattice stretched by an amount ΔL as a force F is applied to both ends, spread out uniformly across the cross-sectional area of the faces A.

This situation is portrayed in figure 9.19.

Each spring in the chain is stretched by *this force* between the atoms on the ends, so that:

$$F_{\text{chain}} = -\frac{Fa^2}{A} = -k_{\text{eff}}\Delta x \tag{9.152}$$

The negative sign just means that the springs are trying to go back to their equilibrium length and hence oppose the applied force.

We multiply this by one in the form $\frac{N_x}{N_x}$ and divide the a^2 over to the right-hand side of the equation and split things up in the following clever way:

$$\frac{F}{A} = -k_{\text{eff}}\frac{N_x}{a^2 N_x}\Delta x = -\frac{k_{\text{eff}}}{a}\frac{N_x \Delta x}{N_x a} \tag{9.153}$$

Note that $aN_x = L$ (again, within one depending on how we count the atoms at the ends of the chain, an error that is negligible if $N_x \gg 1$) and that $N_x\Delta x = \Delta L$. Making these substitutions we get:

$$\frac{F}{A} = -\frac{k_{\text{eff}}}{a}\frac{\Delta L}{L} = -Y\frac{\Delta L}{L} \tag{9.154}$$

where we define:

$$Y = \frac{k_{\text{eff}}}{a} \tag{9.155}$$

This (and the further variants we discss below) is one of the most important equations in the mechanics of solids.

We now give English names to the algebraic forms in the last two equations: F/A is called the ***Stress***, $\Delta L/L$ is called the ***Strain***, and Y is defined to be ***Young's Modulus***. With these definitions, the equation above, stated in English, reads thus:

> ***Compressive or extensive Stress applied to a solid equals Young's Modulus times the Strain.***

(where we don't mention the minus sign because it is understood that, like a spring, the solid *oppposes* the stress to restore its unstressed length).

Here is a short list of things you should know about this equation.

- Stress (F/A) has units of ***pressure***. However, it is ***not*** generally a pressure! Note that in the derivations above, I explicitly *stretched* the material with an *extensive* force rather than *compressed* that material with a *compressive* force pointing towards the material on both ends instead of away. The latter one might be forgiven for calling it a pressure, but the former certainly is not! In both cases, though, the SI units would be pascals, although one could certainly express stress in e.g. bar or even torr.

- Strain ($\Delta L/L$) is ***dimensionless***.

- (Therefore) Young's Modulus (Y) *also* must have (matching) units of pressure: pascals, bar, torr. Again, in no sense should Y be confused with being an *actual* pressure – it is a property of the material based on its linearized microscopic physics!

- In most cases one does not actually compute Y using the definition above (in no small part because computing or measuring the intermolecular potential energy function is extremely difficult). Rather, one reads it out of tables that are worked out empirically! Indeed, one is more likely to work the other way – estimate some of the properties of the intermolecular potential energy from a measurement of Y!

Material	Y (Gpa)
Plastics (various)	1-3
Compact Bone	18
Aluminum	69
Spongy Bone	76
Steel (alloys)	180-200
Tungsten (alloys)	400-600
Diamond	1220

Table 6: A short table of Young's Moduli (in *giga*pascals) for a few interesting materials, to give you an idea of their range. Remember that 10 bar is one megapascal, so these all in the general range of millions of atmospheres

Note well that we can rearrange this equation into "Hooke's Law" as follows:

$$F_x = -\frac{YA}{L}\Delta x = -K\ \Delta x \tag{9.156}$$

which basically states that any given chunk of material *behaves like a spring* when forces are applied to stretch or compress it, with a spring constant K that is ***proportional to the cross sectional area A and inversely proportional to the length L!*** Young's modulus is the material-dependent contribution from the actual geometry and interaction potential energy that holds the material together, but the rest of the dependence is *generic*, and applies to all materials.

This ***scaling*** behavior of the response of solid (or liquid, or gaseous) matter to applied forces is the important take-home conceptual lesson of this section. Substances like bone or steel or wood or nearly anything are *stronger* – respond less to an applied force – as they get *thicker*, and are *weaker* – respond more to an applied force – as they get *longer*. Note well that strength per se is not necessarily directly proportional to Y (or M_s); as we will discuss shortly, the stress at which a material fractures is a separate empirical number and depends on other properties of the material, such as "brittleness". Diamond is extremely hard, for example (Young's modulus in the terapascal range) but it is nevertheless easy to chip because it is brittle.

However, we expect strength to *scale* very much like this as well – longer weaker, thicker stronger. This intuitive understanding can be made quantitatively precise to be sure (and e.g. engineers or orthopedic surgeons will *have* to be quantitatively precise as they craft buildings that won't fall down or artificial hips or bones that mimic natural ones) but is sufficient in and of itself for people to see the world through new eyes, to understand why there are building codes for houses and decks governing the lengths and cross-sectional areas of support beams and joists, why you can't just blow an mouse up to the size of an elephant and still have it able to stand without breaking its own bones, and so on.

We'll explore a few of these applications later, but first let us relatively *quickly* extend the linear extension/compression result to *sideways* forces and understand *shear*.

9.7.4: Shear Forces and the Shear Modulus

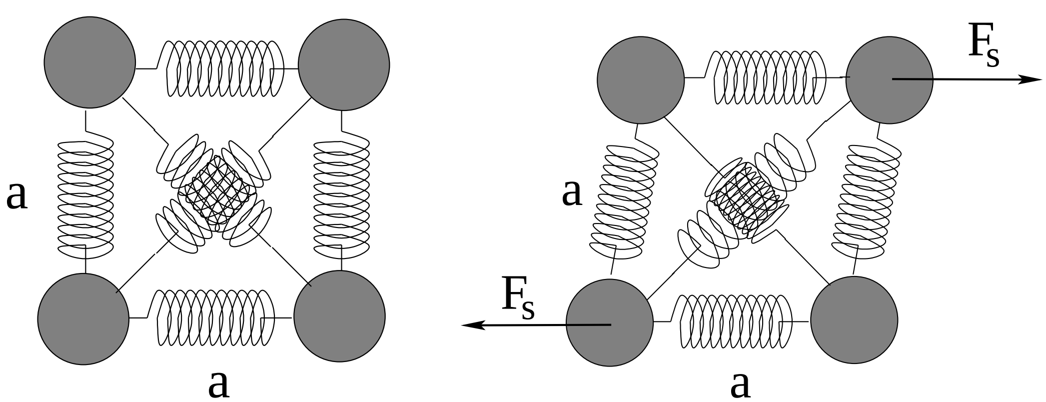


Figure 9.20: A simple model for shear. This model is not particularly correct – in most materials, molecular bonds resist the bending of the stable bond angles at the level of *quantum mechanics* or are more complex (and shear-resistant) lattices.

Let's look at a single 2D "square" of atoms in an imaginary simple cubic lattice like the one pictured in figure 9.18 above. One of the sad things about simple 2D rectangular structures is that they are generally *not* resistant to shear unless angled braces are added – they can flop over from rectangles into parallelograms without altering the lengths of the sides, only the angles at junctions. To allow for that and still have a simple model, in figure 9.20 we portray such a square, but his time we've included some (possibly weaker) molecular bonds that connect opposite corners of the square[239].

In the left-hand figure, no stress is applied. In the right-hand figure, a pair of ***shear stresses*** is applied that form a ***force couple*** (and hence exert a torque!) on the square. We imagine that the bottom of the square is somehow "glued down" so that it cannot move, so instead of rotating, the top of the square is displaced in the direction of the sideways force there, twisting the square into a parallelogram and stretching and compressing the crosswise bonds in the middle as shown.

Now we must imagine an entire block of solid material, "glued" to something along its top and bottom surfaces, with two opposed shear forces applied across the entire area of those surfaces. Internally, all of the molecular bonds will be tipped over as portrayed for our single "atomic" square above, resulting in a net displacement of the top surface (in the direction of the force applied there) to the bottom surface.

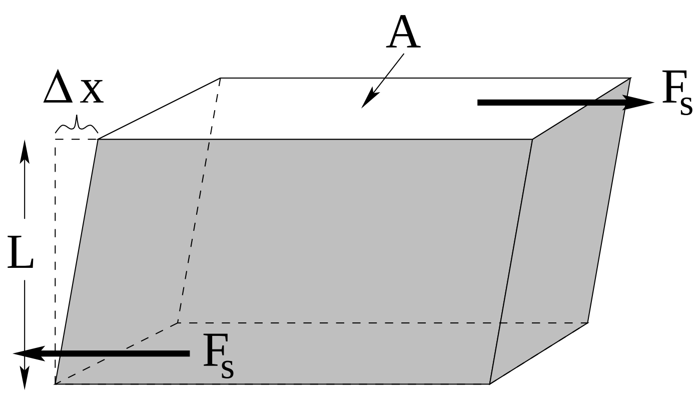


Figure 9.21: A rectangular block of material with *unstressed* dimensions L in height and cross sectional area A. When a shear stress is applied to this block (within the linear response regime) we expect it to deform as shown.

This situation is portrayed in figure 9.21, which also shows the relevant dimensions. The block length (height) we will continue to call L, as this rule will usually be applied to e.g. bones or long support beams, where A is the bone *cross-sectional area* and L its actaal length. Δx is the displacement of the top surface in the direction of the shear force F_s relative to the bottom (which we imagine to be "fixed" and hence is being pulled the other way by an equal and opposite force). The result is that the entire material tips over by a (small) angle θ relative to its unstressed orientation.

A fully microscopic derivation of the expected response is still quite possible, but (obviously) will not be quite as simple as that for Young's modulus above. For that reason we will content

[239]In the most common regular atomic lattices found in nature (things like "face centered cubic" or "body centered cubic") the nearest neighbor bonds alone are sufficient to resist shear. Also, as students of chemistry know, molecular bond angles are typically "fixed" by quantum mechanics (one of the factors that determines the specific geometry of the solid lattice favored by any particular material) and resist shear (bending of these angles) even without "crossbraces".

ourselves with a simple heuristic extension of the *scaling* ideas we learned in the previous section.

The displacement layer to layer required to oppose a given shear force is going to be inversely proportional to the thickness of the material L, because the longer the material is, the more layers there are to split the restoring force among. So we expect the shear force F_s required to produce some given displacement Δx to be proportional to $1/L$. The work we do as the material bends comes primarily from the stretching of the cross-bracing bonds across a single atomic square. Every atom on the top surface has a set of these bonds cross-connecting to the atoms in the direction of the force F_s, and the number of these atoms is proportional to the cross-sectional area A of the material at the shear surface on the top and bottom. The more bonds, the stronger force the force F_s will need to be for the displacement Δx.

We wrap up all of our ignorance of the *specific* microscopic structures that lead to this expected scaling in the linear regime into a modulus that we can more easily *measure in the lab* than *compute from first principles* even as crudely as we did for Young's Modulus, and assemble the result into:

$$F_s = -\frac{M_s A}{L}\Delta x = -K_s \Delta x \tag{9.157}$$

where M_s is the ***shear modulus*** – the equivalent of Young's modulus for shear forces.

Once again we can put this in a form involving *shear stress* (defined to be: F_s/A) and *shear strain* (defined to be $\Delta x/L$):

$$\frac{F_s}{A} = -M_s \frac{\Delta x}{L} \tag{9.158}$$

In words:

Shear Stress applied to a solid equals the Shear Modulus times the Strain.

where note well the minus sign that as always means that the reaction force *opposes* the applied shear force, trying to make Δx *smaller in magnitude* no matter what direction the material is being bent.

As before, $\Delta x/L$ is dimensionless, the units of F_s/A are pressure units (pascals, bar, torr) and hence the units of M_s must be pressure as well. However the shear stress is ***in absolutely no sense an actual pressure!*** The force is being applied *parallel to the surface* and not perpendicular to it!

Shear stress is how springs really work! A "spring" is just a very long piece of material wrapped around into a coil so that when the ends are pulled, the coils *tip* through some small angle in a continuous way, producing shear stress spread out across the entire coil and lengthening the coil itself! So it is not an *accident* that we discover Hooke's Law buried within the discussion in this section – this is a (slightly handwaving) ***microscopic derivation*** of Hooke's Law:

$$F_x = -\frac{M_s A}{L}\Delta x = -k\ \Delta x \tag{9.159}$$

where:

$$k = \frac{M_s A}{L} \tag{9.160}$$

and A is the cross-sectional area of the spring material and L is the entire length of the coiled wire. From this, we can predict how the strength of a spring made of some specific material will vary according to the length of the wire in the coil L and its thickness A. We can also see our addition rules buried in this expression – two identical springs in series are equivalent to one spring twice as long and $k \Rightarrow k/2$; and two identical springs in parallel are equivalent to one spring made with wire with twice the cross sectional area and $k \Rightarrow 2k$! Pretty cool!

9.7.5: Deformation and Fracture

What happens when one stretches or compresses or shears a material by an amount (per bond) that is *not* small? Here is a verbal description of what we might expect based on the microscopic model above and our own experience.

For a range of (relatively small) stresses, the strain remains a ***linear response*** – proportional to the stress, acting in a direction opposed to the stress. As one increases the stress, then, the first thing that happens is that the response becomes *non*-linear. Each additional increment of stress produces *more* than a linear increment of strain when stretching, and quite possibly *less* than a linear increment of strain when compressing, as one can see from the graph of a typical interatomic or intermolecular force in figure 9.17 above. Materials tend to resist compression better than extension because one can always pull bonded atoms *apart*. However, one cannot cause two bonded atoms to *interpenetrate* with *any* reasonable force – they are *very strongly repulsive* when their electron clouds start to overlap but only weakly attractive as the electron clouds are puleed apart.

A second interesting point is that since the total amount of material is conserved, stretching a system at some point starts to make it *thinner* while compressing it makes it *thicker* – this sort of deformation in one direction in response to stresses in another direction is best described by *tensors* and is generally beyond the scope of this course, although the shear modulus is a simple example of one small part of a tensor response.

At some point, one adds enough energy to the bonds (doing *work* as one e.g. stretches the material) that atoms or molecules start to *dislocate* – leave their normal place in the lattice or structure and migrate someplace else. This leaves behind a ***defect*** – a missing atom or molecule in the structure – and correspondingly weakens the entire structure. This migration tends to be *permanent* – even if one releases the stress on the material, it does not return to its original state but remains stretched, compressed, or bent out of place. This region is one of ***permanent deformation*** of the material, as when one bends a paper clip.

Finally, if one continues to ratchet up the stress, at some point the number of defects introduced reaches a critical point and each new defect produced weakens the material enough that another defect is produced without further increase in stress, as the number of bonds over which the stress is distributed decreases with each new defect. Also, the thinning of material on stretching actually decreases the cross-sectional area, which makes it stretch even more in a positive feedback loop. The number of defects "explodes", creating a *fracture* and the material *breaks* (or tears, or crumbles – it comes irreversibly apart).

These behaviors – and the critical points where the behavior changes – are summarized on the graph in figure 9.22.

A material's strength is not (as noted earlier!) just characterized by its various elastic moduli. Elastic moduli describe only the simple linear response regime. Furthermore, the

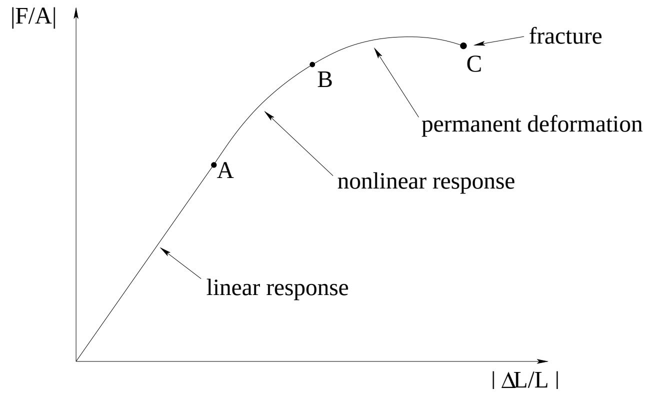


Figure 9.22: The slope of the linear response part of the curve is Young's Modulus (note well the absolute magnitude signs). A marks the point where nonlinear response is first apparent. B marks the boundary where dislocations occur and permanent deformation of the material begins. C marks the point where there is a chain reaction – each defect produced produces on average at least one more defect at the same stress, so that the material "instantly" fractures .

moduli themselves depend on how far the atoms or molecules of the material are away from equilibrium *without* stress of any sort, because the material has a ***temperature*** (basically, a mechanical energy per atom that is *larger* than the minimum associated with sitting at the equilibrium point, at rest). This causes materials to (usually, but not quite always!) *expand* when they are heated by effectively "trapping" a thermal stress inside the material itself.

In thermal equilibrium, each atom is *already* oscillating back and forth around its equilibrium position and temperature increases alone are sufficient to drive hhe system through the same series of states – linear and nonlinear response states where the heated material it can cool back to the original structure, nonlinear response that introduces defects so that the material "starts to melt" and doesn't precisely come back to its original state, followed by *melting* instead of *fracture* when the energy per bond no longer keeps atoms localized to a lattice or structure at all.

When one mixes heating and stress, one can get many different ranges of physical behavior – stressing a system heats it (try bending a paper clip back and forth until it breaks, and it can get hot enough to burn your fingers where it bends). Heating or cooling a system can cause a stress – water is one of the important exceptions to the "expands when heated rule in the vicinity of its freezing point – it actually *[*expands from its greatest density at 4°C to the extent that ice has a density of around 920 kg/m^3 compared to water at 1000 kg/m^2. Water frozen in a confined space exerts and *enormous stress* on the walls confining it. On the other side of four degrees Centigrade, water *heated* in a confined space will exert a large stress on it (even before the water boils). Many, many design decisions in engineering rely on being aware of this coupling between temperature, stress, and strain. Bridges, roadways, siding boards on houses, and much more have to be installed with "expansion joints" or gaps lest the high stresses associated with thermal expansion in a confined space cause structural deformation or failure or the design!

One last qualitative measure of interest when discussing elasticity and strength is *brittleness* or *toughness* – opposing measures of how likely a material is to *bend* (or elastically deform) versus *fracture* (or inelastically deform) when stresses are applied. Some materials, such as steel, are very tough (or not very brittle) – they do not easily deform or fracture and have very large elastic moduli. Others, like diamond, can be very hard indeed (have very large elastic moduli) but are *easy to fracture* – they are *brittle*.

Human bone as a material that has evolved for the specific purpose of structural support varies tremendously in the range of its brittleness with the lifetime of the human involved, with genetic factors, whhh dietary factors, and with the history of the person involved. Yonng people (on average) have bones that bend easily (lower elastic moduli) but aren't very brittle so they don't always break when stressed. Old people have bones that are inflexible but become progressively more brittle as they decalcify with age or disease. Normal adults tend to fall in between – bones that are not as flexible but that are also not particularly brittle.

In all cases *all things being equal* thick bones are stronger than thin bones, long bones are weaker than short bones. This section should have given you a good chance of understanding at least semi-quantitatively how bone strength varies and can be described with a few empirical parameters that can be connected (with a fair bit of work) all the way back to the intermolecular bonds within the bone itself and its physical structure.

9.8: Human Bone

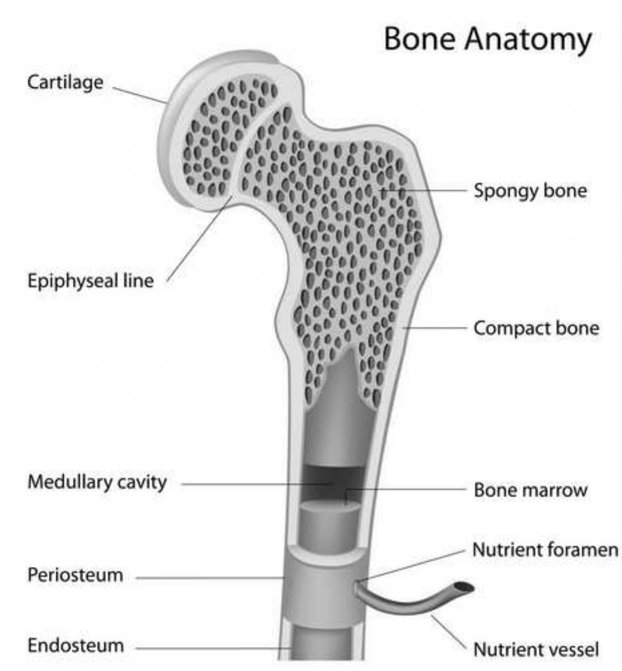


Figure 9.23: This figure illustrates the principle anatomical features of bone.

The bone itself is a composite material made up of a mix of living and dead cells embedded in a mineralized organic matrix. It has significant tensor structure – looking somewhat like a random honeycomb structure in cross section but with a laminated microstructure along the length of the bone. Its anatomy is illustrated in figure 9.23.

Bone is layered from the outside in. The very outer hard layer of a bone is called is ***pe-***

riosteum[240]. In between is compact bone, or ***osteon*** that gives bone much of its strength. Nutrients flow into living bone tissue through holes in the bone called ***foramen***, and are distributed up and down through the osteon through ***haversian canals*** (not shown) that are basically tubes through the bone for blood vessels that run *along* the bone's length to perfuse it. The periosteum and osteon make up roughly 80% of the mass of a typical long bone.

Inside the osteon is softer inner bone called ***endosteum***[241]. The inner bone is made up of a mix of different kinds of bone and other tissue that include spongy bone called ***trabeculae*** and bone marrow (where blood cells are stored and formed). It has only 20% of the bone's mass, but 90% of the bone's surface area. Much of the spongy bone material is filled with blood, to the point where a good way to characterize the difference between the osteon and the trabeculae is that in the former, bone surrounds blood but in the latter, blood surrounds bone.

The bone matrix itself is made up of a mix of inorganic and organic parts. The inorganic part is formed mostly of calcium hydroxylapatite (a kind of calcium phosphate that is quite rigid). The organic part is collagen, a protein that gives bone its toughness and elasticity in much the same way that tough steels are often a mix of soft iron and hard cementite particles, with the latter contributing hardness and compressive/extensive strength, the latter reducing the brittleness that often accompanies hardness and giving it a broader range of linear response elasticity.

There are two types of microscopically distinct bone. ***Woven bone*** has collagen fibers mixed haphazardly with the inorganic matrix, and is mechanically weak. ***Lamellar bone*** has a regular parallel alignment of collagen fibers into *sheets* (lamellae) that is mechanically strong. The latter give the osteon a laminar/layered structure aligned with the bone axis. Woven bone is an early developmental state of lamellar bone, seen in fetuses developing bones and in adults as the initial soft bone that forms in a healing fracture. It serves as a sort of template for the replacement/formation of lamellar bone.

Bones are typically connected together with surface layers of cartilage at the joints, augmented by tough connective tissue and tendons smoothly integrated into muscles that permit mobile bones to be articulated at the joints. Together, they make an impressive mechanical structure capable of an extraordinary range of motions and activities while still supporting and protecting softer tissue of our organs and circulatory system. Pretty cool!

Bone is quite strong. It fractures under compression at a stress of around 170 MPa in typical human bones, but has a smaller fracture stress under tension at 120 MPa and is relatively ¡I¿easy¡/i¿ to fracture with shear stresses of around 52 MPa. This is why it is "easy" (and common!) to break bones with shear stresses, less common to break them from naked compression or tension – typically other parts of the skeletal structure – tendons or cartilage in the joints – fail before the actual bone does in these situations.

Bone is basically brittle and easy to chip, but does have a significant degree of compressive, tensile, or shear elasticity (represented by e.g. Young's modulus in the linear regime) contributed primarily by collagen in the bone tissue. Younger humans still have relatively elastic bones; as one ages one's bones become first harder and tougher, and then (as repair

[240] Latin for "outer bone". But it sounds much cooler in Latin, doesn't it?

[241] Latin for – wait for it – "inner bone".

mechanisms break down with age) weaker and more brittle.

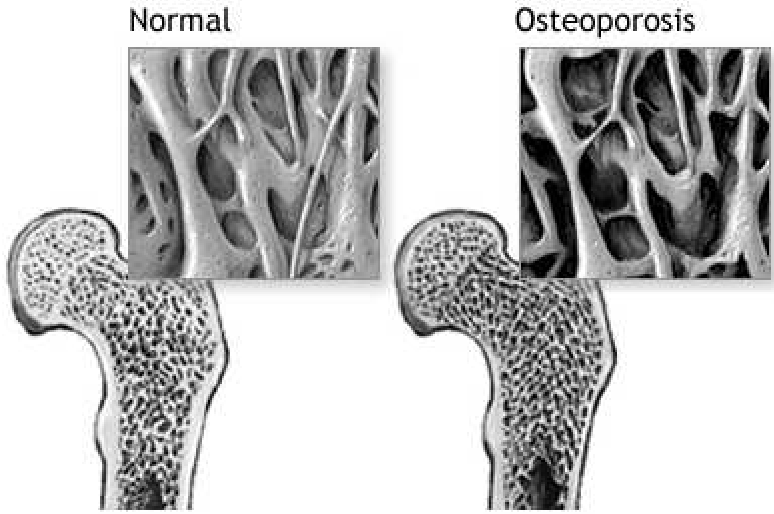


Figure 9.24: Illustration of the alteration of bone tissue accompanying osteoporosis.

Figure 9.24 above shows the changes in bone associated with ***osteoporosis***, the gradual thinning of the bone matrix as the skeleton starts to decalcify. This process is associated with age, especially in post-menopausal women, but it can also occur in association with e.g. corticosteroid therapy, cancer, or other diseases or conditions such as Paget's disease in younger adults.

This process significantly weakens the bones of those afflicted to the point where the static shear stresses associated with muscular articulation (for example, standing up) can break bones. A young person might fall and break their hip where a person with really significant osteoporosis can actually break their hip (from the stress of standing) and then fall. This is not a medical textbook and should not be treated as an authoritative guide to the practice of medicine (but rather, as a basis for understanding what one might by trying to learn *in* a directed study of medicine), but with that said, osteoporosis can be treated to some extent by things like hormone replacement therapy in women (it seems to go with the reduction of estrogen that occurs in menopause), calcium supplementation to help slow the loss of calcium, and certain drugs such as *Fosamax* (Alendronate) that reduce calcium loss and increase bone density (but that have risks and side effects).

Example 9.8.1: Scaling of Bones with Animal Size

An interesting biological example of *scaling laws* in physics – and the reason I emphasize the dependence of many physically or physiologically interesting quantities on length and/or area – can be seen in the scaling of animal bones with the size of the animal[242] . Let us consider this.

[242] Wikipedia: http://www.wikipedia.org/wiki/On Being the Right Size. This and many other related arguments were collected by J. B. S. Haldane in an article titled *On Being the Right Size*, published in 1926. Collectively they are referred to as the *Haldane Principle*. However, the original idea (and 3/2 scaling law discussed below) is due to none other than Galileo Galilei!

We have seen above that the scaling of the "spring constant" of a given material that governs its change in length or its transverse displacement under compression, tension, or shear is:

$$k_{\rm eff} = \frac{XA}{L} \tag{9.161}$$

where X is the relevant (compression, tension, shear) modulus. Bone strength, including the point where the bone fractures under stresses of these sorts, might very reasonably be expected to be proportional to this constant and to scale similarly.

The leg bones of a four-legged animal have to be able to support the weight of that animal under compressive stress. This enables us to make the following scaling argument:

- In general, the weight of any animal is roughly proportional to its volume. Most animals are mostly made of water, and have a density close to that of water, so the volume of the animal times the density of water is a decent approximate guess of what its weight should be.
- In general, the volume of an animal (and hence its weight) is proportional to any characteristic length scale that describes the animal *cubed*. Obviously this won't work well if one compares a snake, with one very long length and two very short lengths, to a comparatively round hippopotamus, but it won't be crazy comparing mice to dogs to horses to elephants that all have reasonably similar body proportions. We'll choose the animal height.
- The leg bones of the animal have a strength proportional to the cross-sectional area.

We would like to be able to estimate the thickness of an animal's bone it's known height and from a knowledge of the thickness of *one kind* of a "reference" animal's bone and its height.

Our argument then is: The volume, and hence the weight, of an animal increases like the cube of its characteristic length (e.g. its height H). The strength of its bones that must support this weight goes like the square of the diameter D of those bones. Therefore:

$$H^3 = CD^2 \tag{9.162}$$

where C is some constant of proportionality. Solving for $D(H)$ we get:

$$D = \frac{1}{\sqrt{C}} H^{3/2} \tag{9.163}$$

This simple equation is approximately satisfied, although not *exactly* as given because our model for bones breaking does not reflect shear-driven "buckling" and a related need for muscle to scale, for mammals ranging from small rodents through the mighty elephant. Bone thickness does indeed increase nonlinearly with respect to body size.

Homework for Week 9

Problem 1.

Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

Problem 2.

Harmonic oscillation is conceptually very important because (as has been remarked in class) many things that are *stable* will oscillate more or less harmonically if perturbed a small distance away from their stable equilibrium point. Draw an ***energy diagram***, a ***graph*** of a "generic" interaction potential energy with a ***stable equilibrium point*** and explain in words and equations where, and why, this should be.

Problem 3.

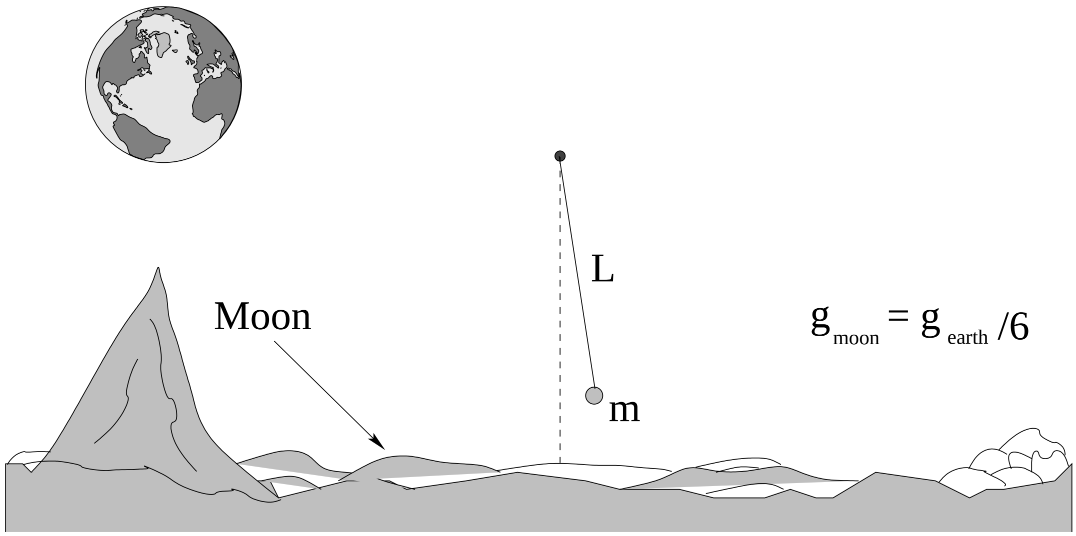
Roman soldiers (like soldiers the world over even today) marched in step at a constant frequency – except when crossing wooden bridges, when they broke their march and walked over with random pacing. Why? What might have happened (and in fact has happened, repeatedly, over historical time) if they marched across with a collective periodic step?

Note that I'm not making this up! You should likely visit:

Wikipedia: http://www.wikipedia.org/wiki/Angers Bridge

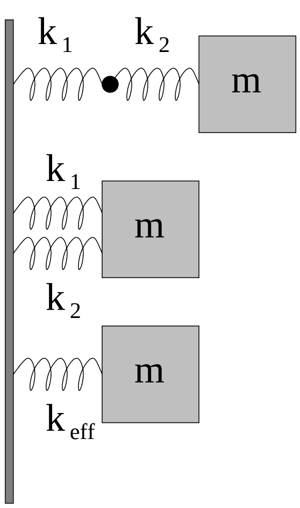
to learn of a famouse case within the last two centuries where failure to break cadence *enough* caused a bridge to collapse, killing over 200 citizens and soldiers!

Problem 4.



A pendulum with a string of length L supporting a mass m on the earth has a certain period T. A physicist on the moon, where the acceleration near the surface is around $g/6$, wants to make a pendulum with the same period. What mass m_m and length L_m of string could be used to accomplish this?

Problem 5.



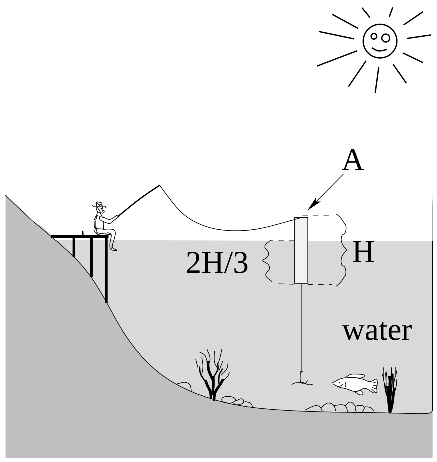
In the figure to the left identical masses m are attached to two springs k_1 and k_2 in "series" – one connected to the other end-to-end – and in "parallel" – the two springs side by side. In the third picture another identical mass m is shown attached to a single spring with constant $k_{\rm eff}$.

a) What should $k_{\rm eff}$ be in terms of k_1 and k_2 for the series combination of springs so that the force on the mass is the same for any given displacement Δx from equilibrium?

b) What should $k_{\rm eff}$ be in terms of k_1 and k_2 for the parallel combination of springs so that the force on the mass is the same for any given displacement Δx from equilibrium?

Note that this problem illustrates the ***scaling*** that underlies Young's modulus with length (series) or area (parallel). It also gives you a head start on how series and parallel addition of resistors and capacitors works next semester! It's therefore well worth puzzling over.

Problem 6.



rgb is fishing in still water off of the old dock. He is using a ***cylindrical*** bobber with a cross sectional area A and a length H. The bobber is balanced so that it remains vertical in the water. When it is floating ***at equilibrium*** (supporting the weight of the hook and worm dangling underneath) $2H/3$ of its length is submerged in the water. You can neglect the volume of the line, hook and worm (that is, their ***buoyancy*** is negligible), and neglect all drag/damping forces. Answer the following questions in terms of A, H, ρ_w (the density of water), and g:

a) What is the combined mass M of the bobber, hook and worm from the data?

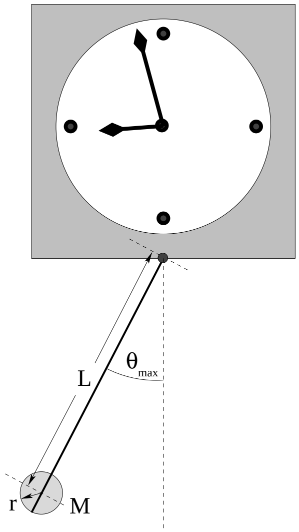
b) A fish gives a tug on the worm and pulls the bobber straight down an ***additional distance*** (from equilibrium) y. What is the ***net*** restoring force on the bobber as a function of y?

c) Use this force and the calculated mass M from a) to write Newton's 2nd Law for the motion of the bobber up and down (in y) and turn it into an equation of motion in standard form for a harmonic oscillator.

d) Assume that the bobber is pulled down the specific distance $y = H/3$ and ***released from rest*** at time $t = 0$. Neglecting the damping effects of the water, write an equation for the displacement of the bobber from its equilibrium depth as a function of time, $y(t)$ (the solution to the equation of motion).

e) With what frequency does the bobber bob? Evaluate your answer for $H = 10$ cm, $A = 1$ cm^2 (reasonable values for a fishing bobber).

Problem 7.



A grandfather clock is constructed with a pendulum that consists of a long, light (assume massless) rod and a small, heavy ***spherical ball*** of mass M and radius $r \ll L$ that can be slid up and down the rod, changing L to "tune" the clock. When the clock is running, the maximum angle the rod makes with the vertical $\theta_{\max}$ is a "small angle". The clock keeps perfect time when the period of undamped oscillation of the pendulum is T_0 seconds.

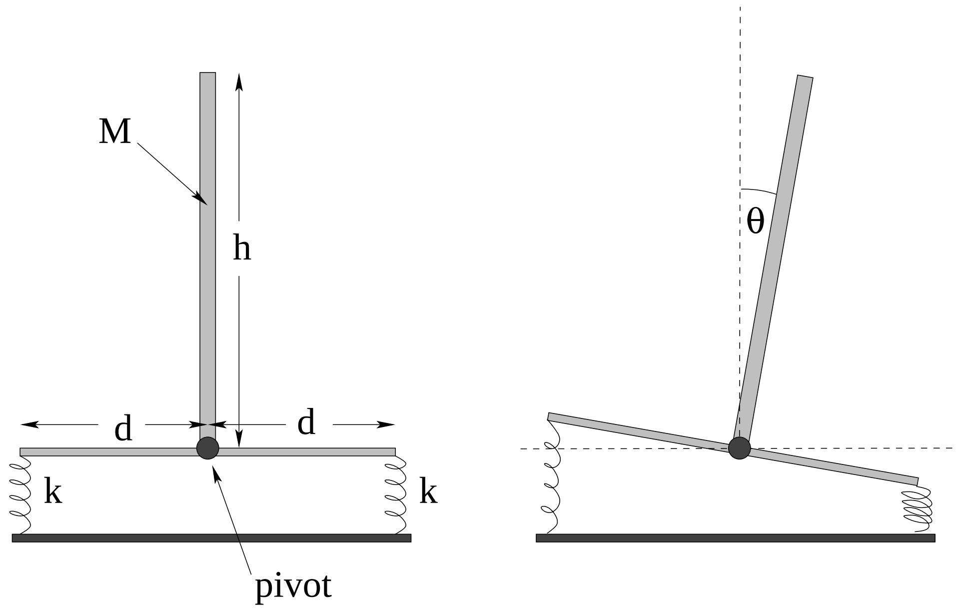
The spherical ball has a moment of inertia of $I_{\rm cm} = \frac{2}{5}Mr^2$ around its own center of mass, recall.

a) *Derive* the equation of motion for the rod when it freely swings and solve for $\theta(t)$ assuming it starts at $\theta_{\max}$ at $t = 0$.

b) At what distance L from the pivot should the mass be set so that the clock keeps correct time (assuming that the small angle approximation is perfectly valid)?

c) Suppose you replace the ball with a new one that has a larger radius $r_{\rm new} > r$. Do you need to move the new ball *up* (to smaller L) or *down* (to larger L) to retune the clock to keep correct time?

Problem 8.



A very light rod of width $2d$ is connected to a heavy table by means of two identical vertical springs with spring constant k attached to the ends as shown. The rod has a pole of mass M and length h welded onto it at right angles to the rod directly above a frictionless axle that acts as a pivot, holding the center of the rod fixed but allowing it to rotate *freely* in the plane of the figure around it. Gravity acts "down", and you can ignore the mass of the rod and friction in the axle in this problem.

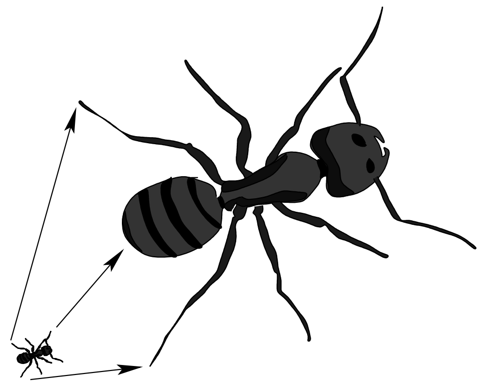
a) On (a copy of) the right-hand figure, indicate the forces acting on the pole and platform when the pole is tipped over at the angle θ. Let positive θ be into the page as shown.

b) Using your answer to a), find the ***total vector torque*** acting on the system (pole and platform) about the hinged bottom of the pole when θ is a *very small angle*, so that $\sin(\theta) \approx \theta$; $\cos(\theta) \approx 1$.

c) Find the ***minimal value*** for the spring constant k of the two springs such that the pole is stable in the vertical position. This means that pulling the pole over to a small angle θ either way generates a *net* torque that ***restores the pole to the vertical***.

d) For $M = 50$ kg, $h = 1.0$ meter, $d = 0.5$ meter, and springs with spring constant $k = 9600$ N/m, *find the angular frequency* $\boldsymbol{\omega}$ with which the pole oscillates about the vertical.

Problem 9.



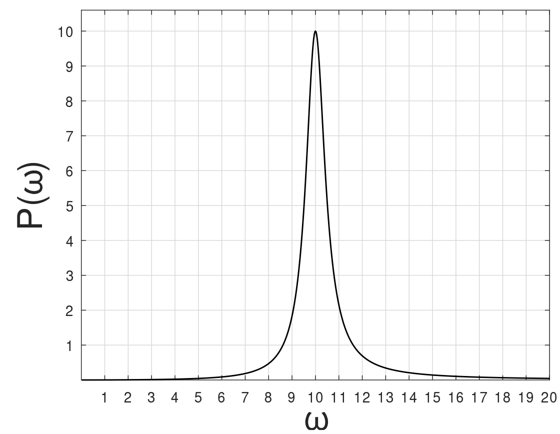
In various science fiction movies, animals that are normally very small are hit with a special ray gun or are given a special chemical in their diet or are irradiated by nuclear test explosions[243] and suddenly end up enormous.

Ants, for example, are noted for being able to carry (around) 20 times their own weight, and are a perennial favorite animal to mutate. The movie "It Came from the Desert" (2017) features perfectly proportional ants that are approximately 10 meters long. A carpenter ant is roughly 1 centimeter in size and we will assign it a mass of 10 milligrams (which may be excessive). We'll assume that this is a super-strong ant and can carry 100 times its own weight, then, it can lift and carry 1 gram or 0.001 kilograms!

Suppose one *did* manage to magnify a carpenter ant perfectly proportionally, so that all of its physical length dimensions were multiplied by 1000. What is the most weight our "monster" carpenter ant could carry? Could it carry its own weight? Is it plausible that it pluck humans from a car (after lifting the car) to snack on?

[243] Yes! Godzilla!

Problem 10.



The curve in the figure above shows the (average) power $P_{\rm avg}(\omega)$ delivered to a damped, driven oscillator with equation of motion:

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = F_0 \cos(\omega t)$$

Recall that the "width" of the curve $\Delta\omega$ is the ***full width at half maximum power***.

Suppose the damping constant b and the mass m are both ***doubled*** while k of the spring and the driving force magnitude F_0 are ***kept unchanged***. What happens to the curve after this change?

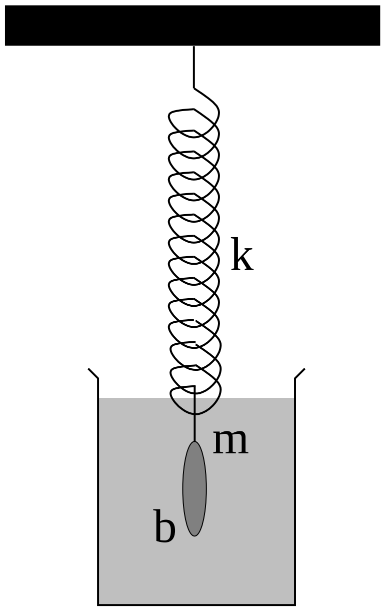
- ☐ The curve becomes narrower.
- ☐ The curve becomes broader.
- ☐ The curve width $\Delta\omega$ does not change;

- ☐ The resonant (peak) frequency is higher.
- ☐ The resonant (peak) frequency is lower.
- ☐ The resonant (peak) frequency does not change.

- ☐ The peak power is higher.
- ☐ The peak power is lower.
- ☐ The peak power does not change.

Be sure to indicate *why* your answers are whatever you select; don't just check the boxes!

Problem 11.



A streamlined mass m is attached to a spring with spring constant k and immersed in a fluid with ***linear*** damping coefficient b. Gravity, if present at all, is irrelevant as shown in class and the textbook. The net force on the mass when it is displaced up by a vertical distance y from equilibrium and moving with upwards with a velocity v_y is thus:

$$F_y = ma_y = -ky - bv_y$$

in one dimension). ***Note Well:*** that the directions of the spring and damping forces with these signs are still correct if y or v_y are negative – they *oppose* the displacement from equilibrium or the velocity respectively, whether that displacement or velocity are themselves positive or negative!

a) Convert Newton's second law for the mass/spring/damping fluid arrangement into the *differential equation of motion* for the system, a "second order, linear, homogeneous, ordinary differential equation" as done in class. (This is a simple rearrangement and division.)

b) Optionally *solve* this equation as done in the textbook and in class, finding in particular the exponential damping rate of the solution (the real part of the exponential time constant) and the shifted frequency ω', assuming that the motion is underdamped, or if you prefer just **put down the solution** derived in class if you plan to *just* memorize this solution instead of learn to derive and understand it.

c) Using your answer for ω' from part b), write down the three criteria for ***underdamped***, ***overdamped***, and ***critically damped*** oscillation.

d) Draw a ***graph*** of $y(t)$ in the case of *weak* (under)damping when $Q = 10$. Use $y_0 = 1$ meter, $\omega_0 = 2\pi$ sec^{-1} (so the period of oscillation is 1 second) to set the vertical and horizontal scale of the graph. It should be at least reasonably quantitatively accurate!

Advanced Problem 12.

In the textbook (and quite probably, lecture) we derived the following expression for the average power delivered to the oscillator in steady state:

$$P_{\rm avg}(\omega) = \frac{F_0^2\omega^2 b}{2m^2\left\{(\omega_0^2-\omega^2)^2+(b\omega/m)^2\right\}}$$

where $\omega_0 = \sqrt{k/m}$ is the angular frequency of the *undamped, undriven* mass on the spring. Do the following:

a) Determine the *maximum* power delivered at the resonant frequency when $\omega = \omega_0$.

b) From this, equate the average power formula to *half* this (algebraic) value. This formula can be used to determine the (two) specific values $\omega_\pm$ at which the power is half-maximum.

c) Solve for these values. The expression (note well) is *quartic*, not really *quadratic*, but if you assume weak damping, then both of $\omega_\pm \approx \omega_0$ and with a bit of insight you will only have to solve for two roots, not four.

d) Use the result to show that the full width at half-maximum power can be written as:

$$\Delta\omega = \omega_+ - \omega_- \approx \frac{b}{m}$$

e) Finally, use this to show that the *same* quality factor we derived for a damped *undriven* harmonic oscillator is also equal to:

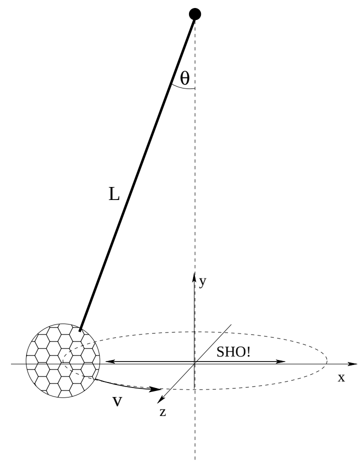
$$Q = \frac{\omega_0}{\Delta\omega}$$

for the damped *driven* harmonic oscillator.

You may find the following factorization useful when $\omega \approx \omega_0$:

$$\omega^2 - \omega_0^2 = (\omega-\omega_0)(\omega+\omega_0) \approx (\omega-\omega_0)\times 2\omega_0$$

Advanced Problem 13.



Let's return to the tether ball homework problem from Chapter/Week 1. Recall that the tether ball of mass m is suspended by a rope of length L from the top of a pole, moving in a circle of radius $r = L\sin(\theta) < L$ at a constant (angular or linear) speed, found to be:

$$v = \sqrt{gL\sin\theta\tan\theta} \quad \Leftrightarrow \quad \Omega = \sqrt{\frac{g}{L\cos\theta}}$$

around the pole. I've added a coordinate frame centered on the circle of motion in the figure above.

Since it moves at a constant speed in a circle in the x-z plane, it is clear that its vector position in this frame can be written something like:

$$\vec{\boldsymbol{x}}(t) = L\sin\theta\cos(\Omega t)\hat{\boldsymbol{x}} + L\sin\theta\sin(\Omega t)\hat{\boldsymbol{z}}$$

(where I'm not worrying about whether it circulates clockwise or counterclockwise viewed from above).

Apparently, the *independent* motion in either x or z is *perfect simple harmonic motion!*

a) What does this astonishing fact tell you about the functional form of the component of $\vec{\boldsymbol{T}}$ (the tension *vector*) in (say) the x-direction? Try to answer this conceptually!

b) *Prove* that your conjecture from a) is correct.

Note well that this is a way of making a *perfect* harmonic oscillator (of sorts) out of a mass and string that works at angles that are *not* "small".

FINAL EXAM IS COMING!

If you have leftover time or energy, continue studying/reviewing for the final exam! Only three “weeks” (chapters) to go in this textbook, finals are coming apace!

Week 10: The Wave Equation

1.17: Wave Summary

- **Wave Equation**

$$\frac{d^2y}{dt^2} - v^2\frac{d^2y}{dx^2} = 0 \tag{10.1}$$

where for waves on a string:

$$v = \pm\sqrt{\frac{T}{\mu}} \tag{10.2}$$

- **Superposition Principle**

$$y(x,t) = Ay_1(x,t) + By_2(x,t) \tag{10.3}$$

(sum of solutions is solution). Leads to interference, standing waves.

- **Travelling Wave Pulse**

$$y(x,t) = f(x \pm vt) \tag{10.4}$$

where $f(u)$ is an arbitrary functional shape or pulse

- **Harmonic Travelling Waves**

$$y(x,t) = y_0 \sin(kx \pm \omega t). \tag{10.5}$$

where frequency f, wavelength λ, wave number $k = 2\pi/\lambda$ and angular frequency ω are related to v by:

$$v = f\lambda = \frac{\omega}{k} \tag{10.6}$$

- **Stationary Harmonic Waves**

$$y(x,t) = y_0 \cos(kx + \delta)\cos(\omega t + \phi) \tag{10.7}$$

where one can select k and ω so that waves on a string of length L satisfy fixed or free boundary conditions.

- **Energy (of wave on string)**

$$E_{\mathrm{tot}} = \frac{1}{2}\mu\omega^2 A^2 \lambda \tag{10.8}$$

is the total energy in a wavelength of a travelling harmonic wave. The wave transports the power

$$P = \frac{E}{T} = \frac{1}{2}\mu\omega^2 A^2 \lambda f = \frac{1}{2}\mu\omega^2 A^2 v \tag{10.9}$$

past any point on the string.

- **Reflection/Transmission of Waves**

  a) Light string (medium) to heavy string (medium): Transmitted pulse right side up, reflected pulse inverted. (A fixed string boundary is the limit of attaching to an "infinitely heavy string").

  b) Heavy string to light string: Transmitted pulse right side up, reflected pulse right side up. (a free string boundary is the limit of attaching to a "massless string").

- **Harmonic Wave Relations to "Just Know", in "The Box"**

$$\boxed{\begin{array}{c} k = \dfrac{2\pi}{\mathcal{T}} \qquad k = \dfrac{2\pi}{\lambda} \qquad kv = \omega \qquad f = \dfrac{1}{\mathcal{T}} \\ v = \dfrac{\omega}{k} = \dfrac{2\pi/\mathcal{T}}{2\pi/\lambda} = \dfrac{\lambda}{\mathcal{T}} = f\lambda = \sqrt{\dfrac{T}{\mu}} \end{array}}$$

10.1: Waves

We have seen how a particle on a spring that creates a restoring force proportional to its displacement from an equilibrium position oscillates harmonically in time about that equilibrium. What happens if there are *many* particles, *all* connected by tiny "springs" to one another in an extended way? This is a good metaphor for many, many physical systems. Particles in a solid, a liquid, or a gas both attract and repel one another with forces that maintain an average particle spacing. Extended objects under tension or pressure such as strings have components that can exert forces on one another. Even fields (as we shall learn next semester) can interact so that changes in one tiny element of space create changes in a neighboring element of space.

What we observe in all of these cases is that changes in any part of the medium "propagate" to other parts of the medium in a very systematic way. The motion observed in this propagation is called a *wave*. We have all observed waves in our daily lives in many contexts. We have watched water waves propagate away from boats and raindrops make circular waves rippling away from where they splash onto the surface of a calm pond. We have listened to musical sound waves generated by waves created on stretched strings or from tubes driven by air and transmitted invisibly through space by means of radio waves. We read these words by means of light, an electromagnetic wave. In advanced physics classes one learns that all matter is a sort of quantum wave, that indeed *everything* is really a manifestation of waves. *We* are basically very complicated quantum waveforms!

It therefore seems sensible to make a first pass at understanding waves and how they work in general, so that we can learn and understand more in future classes that go into detail.

The concept of a wave is simple – it is an *extended structure* that oscillates in both *space* and in *time*. We will study two kinds of waves in this particular Mechanics and Applications textbook:

- **Transverse Waves** (e.g. waves on a string). The displacement of particles in a transverse wave is *perpendicular* to the direction of the wave itself.

- **Longitudinal Waves** (e.g. sound waves). The displacement of particles in a longitudinal wave is in the same direction that the wave propagates in.

Some waves, for example water waves, are simultaneously longitudinal and transverse. Transverse waves are probably the most important waves to understand for the future; light is a transverse wave. We will therefore start by studying transverse waves in perhaps the simplest context where we can easily go from Newton's Second Law to the One Dimensional Wave (differential) Equation: waves on a stretched string.

10.2: Waves on a String

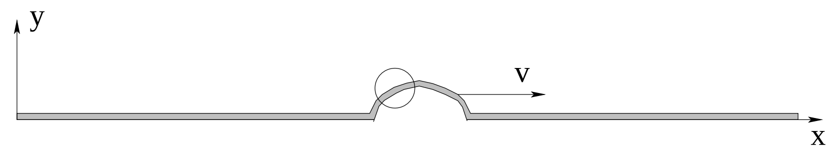


Figure 10.1: A uniform string is plucked or shaken so as to produce a *wave pulse* that travels at a speed v to the right. The circled region is examined in more detail in the next figure.

Suppose we have a uniform *stretchable* string (such as a guitar string) that is pulled at the ends so that it is under tension T. The string is characterized by its mass per unit length μ – thick guitar strings have more mass per unit length than thin ones but little else. It is fairly harmless at this point to imagine that the string is fixed to pegs at the ends that maintain the tension. Our experience of such things leads us to expect that the stretched string will form an almost perfectly straight line unless we pluck it or otherwise bend some other shape upon it. We will impose coordinates upon this string such that x runs along its (undisplaced) equilibrium position and y describes the vertical displacement of any given bit of the string.

Now imagine that we have plucked the string somewhere between the end points so that it is displaced in the y-direction from its equilibrium (straight) stretched position and has some curved shape, as portrayed in figure 10.1. The tension T, recall, acts all along the string, but because the string is *curved* the force exerted on any small bit of string does not balance. This inspires us to try to write Newton's Second Law not for the entire string itself, but for just a tiny bit of string, indeed a *differential* bit of string.

We thus zoom in on just a small chunk of the string in the region where we have stretched or shaken a *wave pulse* as illustrated by the figure above. If we blow this small segment up, perhaps we can find a way to write the unbalanced forces out in a way we can deal with algebraically.

This is shown in figure 10.2, where I've indicated a short segment/chunk of the string of length Δx by cross-hatching it. We would like to write Newton's 2nd law for that chunk. As you can see, the same *magnitude* of tension T pulls on both ends of the chunk, but the tension pulls in slightly different directions, tangent to the string at the end points.

If θ is small, the components of the tension in the x-direction:

$$\begin{aligned} F_{1x} &= -T\cos(\theta_1) \approx -T \\ F_{2x} &= T\cos(\theta_1) \approx T \end{aligned} \tag{10.10}$$

are nearly equal and in opposite directions and hence nearly perfectly cancel (where we have used the small angle approximation to the Taylor series for cosine:

$$\cos(\theta) = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots \approx 1 \tag{10.11}$$

for small $\theta \ll 1$).

Each bit of string therefore moves more or less *straight up and down*, and a useful solution is described by $y(x,t)$, the y **displacement** of the string at position x and time t. The permitted

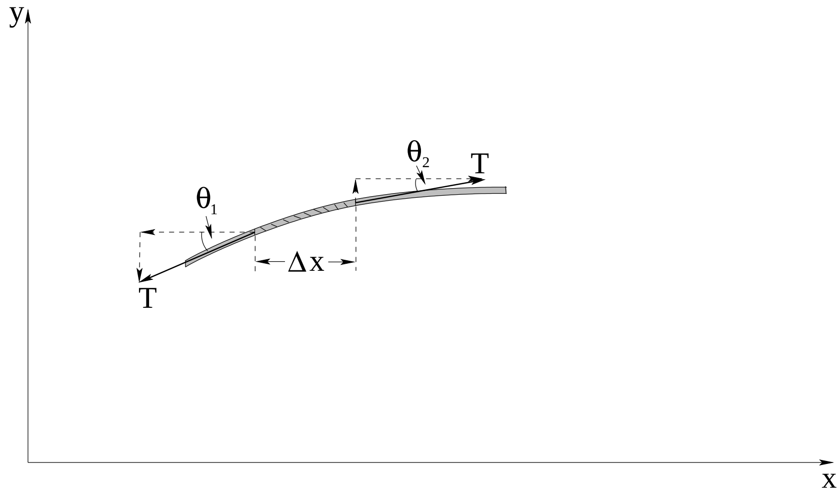


Figure 10.2: The forces exerted on a small chunk of the string by the tension T in the string. Note that we neglect gravity in this treatment, assuming that it is much too small to bend the stretched string significantly (as is indeed the case, for a taut guitar string).

solutions must be *continuous* if the string does not break. It is worth noting that not *all* waves involving moving up and down only – this initial example is called a *transverse* wave, but other kinds of waves exist. Sound waves, for example, are *longitudinal* compression waves, with the motion back and forth along the direction of motion and not at right angles to it. Surface waves on water are a mixture, they involve the water *both* moving up and down *and* back and forth. But for the moment we'll stick with these very simple transverse waves on a string because they already exhibit most of the properties of far more general waves you might learn more about later on.

In the y-direction, we find write the force law by considering the total y-components of the sum of the force exerted by the tension on the ends of the chunk:

$$\Delta F_{y,tot} = F_{1y} + F_{2y} = -T\sin(\theta_1) + T\sin(\theta_2) = \Delta m a_y = \mu \Delta x a_y \tag{10.12}$$

Here ΔF_y is not the "change in the force" but rather the total force acting on the chunk of length Δx. These two quantities will obviously scale together.

We make the small angle approximation: $\sin(\theta) \approx \tan(\theta) \approx \theta$:

$$\Delta F_y = T\tan(\theta_2) - T\tan(\theta_1) = T(\tan(\theta_2) - \tan(\theta_1)) = \mu \Delta x a_y, \tag{10.13}$$

Next, we note that $\tan(\theta) = \frac{dy}{dx}$ (the slope of the string is the tangent of the angle the string makes with the horizon) and divide out the $\mu\Delta x$ to isolate $a_y = d^2y/dt^2$. We end up with the following expression for what might be called *Newton's Second Law per unit length* of the string:

$$\frac{\Delta F_y}{\Delta x} = T\frac{\Delta(\frac{dy}{dx})}{\Delta x} = \mu\frac{d^2y}{dt^2} \tag{10.14}$$

In the limit that $\Delta x \to 0$, this becomes:

$$f_y = \frac{dF_y}{dx} = T\frac{d^2y}{dx^2} = \mu\frac{d^2y}{dt^2} \tag{10.15}$$

where we might consider $f_y = dF_y/dx$ the *force density* (force per unit length) acting at a point on the string. Finally, we rearrange to get:

$$\frac{d^2y}{dt^2} - \frac{T}{\mu}\frac{d^2y}{dx^2} = 0 \tag{10.16}$$

The quantity $\frac{T}{\mu}$ has to have units of $\frac{L^2}{t^2}$ which is a velocity squared.

We therefore formulate this as the ***one dimensional wave equation***:

$$\frac{d^2y}{dt^2} - v^2\frac{d^2y}{dx^2} = 0 \tag{10.17}$$

where

$$v = \pm\sqrt{\frac{T}{\mu}} \tag{10.18}$$

are the allowed velocities of the wave on the string. This is a **second order, linear, homogeneous, ordinary differential equation**[244] and has (as one might imagine) well known and well understood solutions.

We'll be referring to the "one dimensional wave equation" a *lot* below, and I'm lazy. I will therefore most often use **1DWE** in most of our discussion to refer to this particular differential equation just as we used SHOE in the last chapter. As was also the case in the last chapter, the actual quantity appearing in the 1DWE[245] need not be the transverse displacement of a physical stretched string carrying a wave that satisfies the small-angle approximation – it can refer to longitudinal displacement or pressure (next chapter) or a component of e.g. the electric field of a light wave (next semester/course). In the meantime, in *this* chapter the 1DWE[246] *will* refer only to transverse waves on a string.

Note well: As the tension in the string increases, so does the wave velocity. As the mass density of the string increases, the wave velocity decreases. This makes *physical sense*. As tension goes up the restoring force is greater. As mass density goes up one accelerates less for a given tension. It is useful to remember that the 1DWE is *just Newton's Second Law in disguise* for a stretched string, and Newtonian concepts of mass, force, acceleration and energy still apply!

10.2.1: An Important Property of Waves: Superposition

The 1DWE for waves on a string is *linear* – which just means that y appears only to the first power – and *homogeneous* (for now) which means that there is a *zero* on the right hand side instead of (say) a source function of some sort. It is easy to show that if one function $y_1(x,t)$ solves the 1DWE (or, really, any linear homogeneous ODE) and a second function $y_2(x,t)$ (independent of y_1) *also* solves the 1DWE, then:

$$y(x,t) = Ay_1(x,t) + By_2(x,t) \tag{10.19}$$

[244] Impress the heck out of your friends when you tell them what *you* did today. "I learned to solve a second order, linear, homogeneous ordinary differential equation. And you? How was basket weaving today?"

[245] See? I did it for real right there...

[246] ...and again!

solves the 1DWE for arbitrary (complex) A and B.

This property of waves is most powerful and sublime. Let's prove it, as it occurs again and again in physics and mathematics and engineering, in many contexts! The proof is pretty trivial – we just apply the (differential) "wave equation" as *operators* from the left to this sum on both sides:

$$\begin{aligned}
\frac{d^2y}{dt^2} - v^2\frac{d^2y}{dx^2} &= \left(\frac{d^2}{dt^2} - v^2\frac{d^2}{dx^2}\right)\Big(Ay_1(x,t) + By_2(t)\Big) \\
&= \left(\frac{d^2Ay_1}{dt^2} - v^2\frac{d^2Ay_1}{dx^2}\right) + \left(\frac{d^2By_2}{dt^2} - v^2\frac{d^2By_2}{dx^2}\right) \\
&= A\left(\frac{d^2y_1}{dt^2} - v^2\frac{d^2y_1}{dx^2}\right) + B\left(\frac{d^2y_2}{dt^2} - v^2\frac{d^2y_2}{dx^2}\right) \\
&= A\times 0 + B\times 0 = 0 \qquad (10.20)
\end{aligned}$$

Almost embarrassingly simple, right? Since y_1 and y_2 independently solve the 1DWE, the stuff in the parentheses all vanishes once I've factored out the *constants* A and B. But simple or not, this is an *enormously powerful conclusion* and one that is the foundation of a huge chunk of STEM[247] knowledge.

Hopefully it is obvious that we don't have to stop at just two solutions (if we happen to have more than two handy). We've really proven that if $y_1, y_2, y_3...$ satisfy the 1DWE (or any *other* linear homogeneous differential equation), then an arbitrary linear superposition of those solutions:

$$y(x,t) = a_1y_1(x,t) + a_2y_2(x,t) + a_3y_3(x,t) + ... = \sum_i a_iy_i(x,t)$$

is also a solution! In a second, we'll prove that there are in fact an *infinite* number[248] of solutions to the 1DWE, although we'll leave this as a forking point for more advanced students to pursue on their own, or in later courses, while we'll continue down the simple/intro path here.

10.3: Solving the IDWE for Traveling Waveforms

In one dimension there are at least three distinct solutions to the wave equation that we are interested in. Two of these solutions *propagate* along the string – energy is *transported* from one place to another by the wave. The third is a *stationary* solution, in the sense that the wave doesn't propagate in one direction or the other (not in the sense that the string doesn't move).

The traveling wave solution is actually almost trivial to deduce, without actually "solving" for much of anything, so we'll work first on deducing it before (eventually) looking at some slightly fancier math that let's us home in on the *harmonic* solutions that are perhaps of the greatest interest.

We start by noting that the 1DWE equates a second derivative in our *time* coordinate t to (the constant v^2 times) a second derivative in our *space* coordinate x. Hmmm, we think.

[247] Science, Technology, Engineering and Mathematics, although I personally would write it as STEMM, Science, Technology, Engineering, Mathematics, *and Medicine* as modern medicine is STEM stuff almost all the way...

[248] *Uncountable* infinite number, for those who care about such things...

Could it be that simple? Could we – no joke – pick almost *any twice continuously differentiable function of a single variable* and turn it into a solution to the 1DWE?

We can! Suppose we write the solution as:

$$y(x,t) = f(u)$$

where u is an *unknown function of* x *and* t, substitute it into the differential equation, and use the chain rule(s) $\frac{d}{dt} = \frac{du}{dt}\frac{d}{du}$ and $\frac{d}{dx} = \frac{du}{dx}\frac{d}{du}$, twice. This is just:

$$\frac{d^2 f}{dt^2} - v^2 \frac{d^2 f}{dx^2} = 0 \tag{10.21}$$

$$\frac{d^2 f}{du^2}\left(\frac{du}{dt}\right)^2 - v^2 \frac{d^2 f}{du^2}\left(\frac{du}{dx}\right)^2 = 0 \tag{10.22}$$

$$\frac{d^2 f}{du^2}\left\{\left(\frac{du}{dt}\right)^2 - v^2 \left(\frac{du}{dx}\right)^2\right\} = 0 \tag{10.23}$$

From this we see that as long as $\frac{d^2 f}{du^2} \neq 0$, the solution *requires* the following relation between u, x and t:

$$\left\{\left(\frac{du}{dt}\right)^2 - v^2 \left(\frac{du}{dx}\right)^2\right\} = 0 \quad \Rightarrow \quad \frac{du}{dt} = \pm v \frac{du}{dx} \tag{10.24}$$

Suppose we let:

$$\frac{du}{dx} = 1 \quad \Rightarrow \quad u = x + (\text{something independent of } x)$$

Why not? It has the right units, after all. Then:

$$\frac{du}{dt} = \pm v \quad \Rightarrow \quad u = (\text{something independent of } t) \pm vt$$

Put them together, we get:

$$u = x \pm vt \tag{10.25}$$

What this tells us is that *any* function $f(u)$ that is at least *twice continuously differentiable in* u[249] – so that its second derivative is not *zero* everywhere or *infinity* anywhere we care about is a solution to the wave equation as long as we write it like:

$$y(x,t) = f(x \pm vt) \tag{10.26}$$

and arrange for f to have the same units/dimensions as y.

There is, however, one ***extremely serious problem*** with the $f(u = x \pm vt)$ solution, one that I have ignored in the discussion so far but that is there nonetheless. $\boldsymbol{u = x - vt}$ ***is not dimensionless!***. This is actually a *serious problem* because, for physics to be *consistent*:

[249]...and satisfies the small angle condition we used to derive the 1DWE for waves on a string in the first place, and is multiplied as necessary by something that gives it the right units, and maybe a couple more conditions, like the next one I discuss, so don't go *completely* crazy here...

> **The argument of any kinematic function that has a power series (e.g. Taylor series) expansion in physics *must be dimensionless!***

My favorite adage for remembering and understanding this is that we cannot add lengths in meters (e.g. x) to liters (e.g. x^3). Or to areas. Or to fifth powers of the length. Hence anytime we see a function that has a power series expansion, such as:

$$e^u = 1 + u + \frac{1}{2!}u^2 + \frac{1}{3!}u^3 + ...$$

we *know* that x is either dimensionless or (if it has units of length) we *know* there is a hidden constant with units of inverse length and value 1 multiplying it!

We've already encountered this – in week 1 homework, for example, and in the SHOE solutions in the last chapter – as well as off and on throughout the various topics in the text, and it tells us that our proposed harmonic solution is *missing a crucial scale factor* that must have the ***units of inverse length***[250]. I'll address this point in context as I develop specific solutions below.

Anyway, our extremely simple analysis above didn't find *one* solution to the 1DWE, it turned up a literal infinity of solutions[251]. *Any* twice differentiable single-variable functional shape of wave – referred to as a "waveform" – created on the string and (as we shall see) propagating to the right or left at speed v is a solution to the wave equation as long as its u argument is scaled appropriately to be dimensionless!

We are going to concentrate on just three "solutions of interest" that illustrate various important properties of waves and solutions to the 1DWE in general. All three can be connected to solutions of this sort, although for one of them it will make more sense to derive solutions independently as they are not traveling waves (so using a traveling waveform doesn't really make sense).

The first of them *is* a traveling waveform, however: a "wave pulse", which we will use to understand a number of things about waves on a string for now, and later for other kinds of waves as well.

10.4: Wave Pulses

If you've ever pulled a garden hose across your yard and hooked it on a toy or tree stump and felt too lazy to put down your end, walk all the way over to unhook at, walk all the way back and continue to drag the hose, you've probably tried to *flip the hose off* from where you stand[252]. That is, you may have moved your end of the hose up and down *rapidly*, and watched a ripple run down from your hand to the obstacle and quite possibly *lift the hose free* for an instant and allow you to pull the hose on by it.

[250] Or, perhaps, it is there but *happens* to have value '1', as it did in the week 1 homework...

[251] Most of which we cannot even write down. We will, therefore, concentrate on only a few important ones where we *can*...

[252] ...and no, this is not a rude reference, although you might have been *tempted* to flip off the recalcitrant hose the other way...

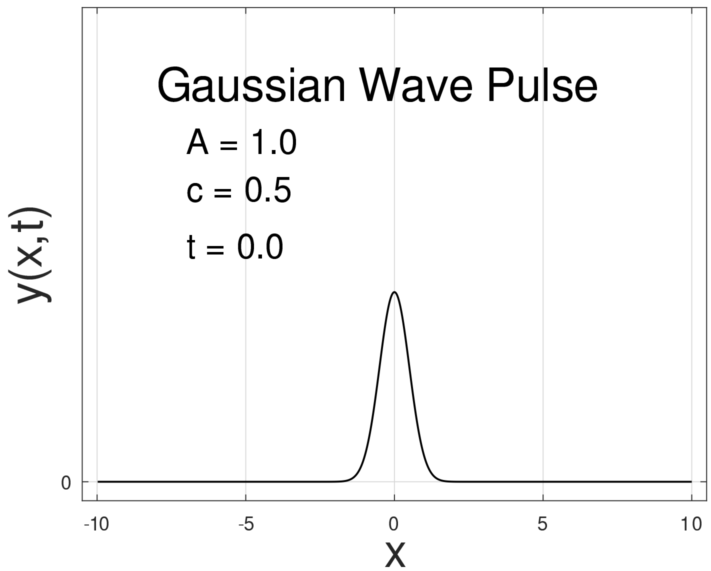


Figure 10.3: A “generic” wave pulse, centered on the origin. I actually plotted a gaussian waveform, with height $A = 1.0$ and width parameter $c = 0.5$, but could have used any other peaked function instead or just drawn a bell-ish curve by hand!

If so (or done something similar with ropes, strings, wires, etc) you have already experimented with ***wave pulse*** solutions to the 1DWE. These are solutions where $f(u)$ is a *relatively narrow, relatively sharply peaked function* that can run down a rope, string, garden hose, wire under at least a bit of tension. Think “Gaussian shape” or “bell curve” if you happen to be familiar with what that is from e.g. a statistics class.

It isn't crucial in *this* course that we actually write down an actual function in actual coordinates for a wave pulse, only that we can visualize one and understand what is *meant* by the term “wave pulse”, but I'd feel guilty if I didn't give you at least *one* such function as an example, so in figure 10.3 I drew an actual gaussian function (class “bell curve” form from e.g. statistics) located (peaked) at the origin at time $t = 0$:

$$y(x,t) = Ae^{-\frac{(x \pm vt)^2}{2c^2}}$$

where $A = 1.0$ and $c = 0.5$, just so you could look at it and perhaps better viualize how it propagates later. Note well, this *particular* function peaks ***where $x \pm vt = 0$!***

There *is* one very useful bit of pedagogy we can pull out of this waveform – recall that we *just* learned/reviewed the requirement that the argument of a candidate $f(u)$ be dimensionless. Well, the gaussian function has $(x \pm vt)^2$ on top of its argument. On the bottom it has a c^2. From this we can see that ***c must have units of length*** and indeed *c sets the scale of the gaussian, making it wider or narrower in whatever units we are using for x!* Since $y(x,t)$ presumably has units of length as well (if it is to describe a transverse wave on a *string*) I made two errors in the previous paragraph. Can you spot them?

I *should* have written $A = 1.0$ *millimeters* (for example) and $c = 0.5$ *centimeters* (for example). Both of these are dimensioned constants and cannot be given just by a number without units!

For now, though, forget the units per se, concentrate on the shapes and how they evolve in time. We can use wave pulses *graphically* to help us understand *just how waves propagate to the left or the right* depending on the ***sign of*** $\pm vt$ in $f(x \pm vt)$, how waves reflect off of string junctions or connections to walls, and much more, all without actually needing to write down a formula for one, but if you are a math, physics, or stats major it shouldn't have done you any harm hurt to see at least the one above written down.

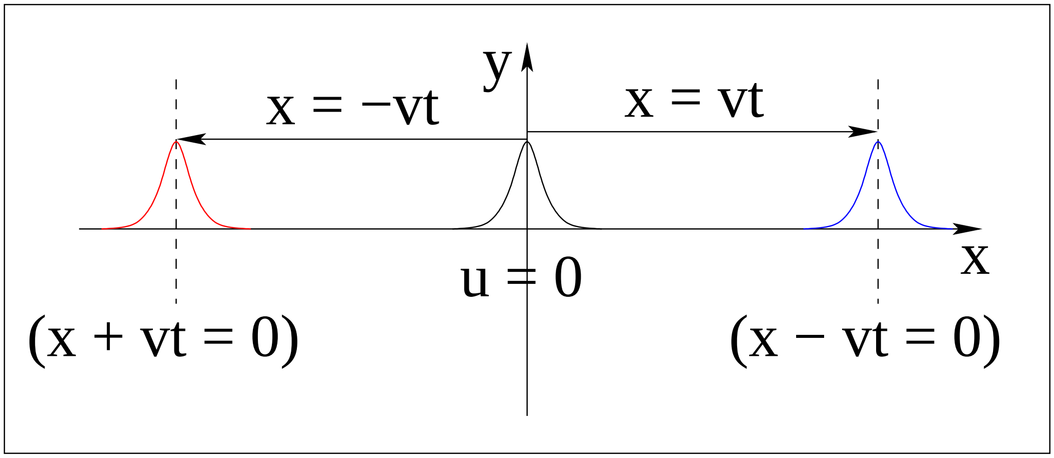


Figure 10.4: Wave pulses like the one illustrated in figure 10.3 above start at the origin (black) at $t = 0$. This one I *did* draw "by hand" with xfig. Not much difference. If this pulse propagates to the *right* to arrive at position $x = vt$ at time t, it looks like the blue one. If it starts at the origin and propagates to the *left* to arriving at $x = -vt$ at time t, it looks like the red one. Pretty simple idea, actually.

Let's graphically figure out the first of these. How *does* the sign of $\pm vt$ determine the direction of the propagation of a travelling wave? Consider figure 10.4 above.

Wow! Note how the graph alone explains it all! The wave pulse is peaked where $x \pm vt = 0$, as noted. If it starts at the origin at $t = 0$ and ends up at the blue position at time t, the peak is now at $x = vt \Leftrightarrow x - vt = 0$! If it starts at the origin at $t = 0$ and ends up at the red position at time t, the peak is now at $x = -vt \Leftrightarrow x + vt = 0$! We conclude that:

$y(x, t) = f(x - vt)$ is a wave form propagating to the ***right!***

$y(x, t) = f(x + vt)$ is a wave form propagating to the ***left!***

10.4.1: Reflection of Wave Pulses

One problem we haven't addressed is that in order for a string to be at tension $\mathcal{T}$, it pretty much *has to be pulled in opposite directions at both ends!* We have to affix the string to *something* – tie it to a nail, hold it with a finger, suspend it on a rod with a loop – and *pull*.

This often imposes *boundary conditions* on the allowed solutions to the 1DWE[253] so that we can't, actually use *any* $f(u)$, we have to restrict ourselves to ones that satisfy the boundary

[253] I warned you not to go all crazy, right? Here we are already going from a huge infinite number of solutions to a smaller infinite number of solutions. Indeed, often from an uncountable infinity to a – sort of – *countable* infinity.

conditions, for example that are always zero at the point where we fastened a string to keep it under tension.

Well, we know that we can tie a string to a wall (fixing that end) and hold the other in our hand and pull to put it under tension. We also know that if we rapidly "shake" our hand up and down, we'll put a wave pulse on the string that then runs away from our hand. Suppose we have done this to put a wave pulse on the string that is incident on the fixed end where it connects to the wall. What will happen when it arrives at that fixed point?

There are several ways of seeing what the answer must be. We haven't discussed work and energy – yet – but clearly our hand did work to put the wave pulse on the string, so it must have some energy associated with it, and it must actually be carrying that energy along *with* the pulse as it propagates! That, in a nutshell, is why waves are so interesting! They are nature's way of transporting energy from one place to another so "interesting" things – like us – can happen!

However, the point where the string is tied to the wall is *fixed*. It does not move! It therefore cannot do any work on the string as the pulse arrives there! Whatever energy was in the pulse can't keep going as there is no more string. It can't just disappear – energy is conserved. It can't even be changed, as the nail the string is tied to does no work on the string.

We conclude (and quickly confirm by observation) that the wave must be *reflected* from the fixed point, reflected in just the right way for the total energy in the wave pulse not to change *and* for the string to never actually move at the point it is tied to the wall.

Deriving the result is actually pretty hard, but we can actually find a solution to the 1DWE that satisfies these two conditions really easily by just thinking for a moment.

Suppose the string were *not* tied to the wall – suppose it was in fact twice as long and the other end was being held by a student just like you. At some pre-agreed upon time, *you* flip a wave pulse *up* and send it zinging off towards the other student. They simultaneously flip an otherwise identical wave pulse *down* and send it towards *you*.

What will happen in the exact center of the string at the point where the string *would* have been tied? Well the two pulses will cross one another, exactly upside down. From the superposition principle, we know that the two pulses will just *add*, and that:

a) The energy that moves to the right of the midpoint from the first pulse will be precisely replaced by the second pulse moving the other way. The total energy on your *half* of the string will not change.

b) The exact midpoint of the string will be a point where the two "mirror image" pulses exactly cancel, so that point will not move.

Now we just need to use our imagination. *We could have fastened the string to a fixed rod right in the middle and nothing would change!* On *your* half of the string, the wave pulse you sent at the wall would reach the rod and reflect from it, *flipping the pulse upside down*. Your partner would see the exact same thing, upside down. We conclude that:

Waveforms reflecting from a fixed point invert.

with *no other change*. They just go back the way that they came in, upside down.

There is one other way we can imagine keeping tension on a string. We can attach the string to a small "massless, frictionless" ring, thread a frictionless rood through it and then pull the rod back at right angles to the string! The string is under tension, but the end is not fixed, it is *completely free* to move up and down the rod with no resistance.

Newton's third law tells us that while the string can freely move up or down the rod as a wave pulse arrives there, the string must *always come into the rod at right angles* because its tension is maintained by a *normal* force from the rod, with no component at all along the rod. Instead of the wave function itself being always zero, as it is at a fixed point on the string, it's *slope* is always zero at the rod. And as before, the rod can do no work on the string because the force it exerts is always at right angles to the motion of the string up and down the rod, so the energy in any waveform incident on the rod must be conserved as the waveform is reflected.

Again our imagination tells us how to build a solution that satisfies these two conditions. We just have to tell our friend to put a pulse on their end of the string *right side up* – the same direction as your own. When the two pulses cross, they will add *up*, not cancel, at the midpoint of the string. Whatever slope your pulse has, the other pulse will have the opposite slope and the resulting wave will have zero slope (but nonzero value) at the midpoint as the two pulses cross.

You will see the "reflected" pulse come back at you *right side up* and otherwise identical to the one you sent in, and be guaranteed as before that:

a) The energy that moves to the right of the midpoint from the first pulse will be precisely replaced by the second pulse moving the other way. The total energy on your *half* of the string will not change.

b) The exact midpoint of the string will be a point where the two "mirror image" pulses exactly add, so that the slope of the string there remains zero.

and:

Waveforms reflecting from a free point remain right side up.

with *no other change*. They just go back the way that they came in, right side up.

In summary, wave pulses *invert* when reflected from a fixed boundary (string fixed at one end) and reflect right side up from a free boundary (string free to slide vertically at one end with no friction). In both cases the connection can maintain tension in the string as the pulse hits it.

When two strings of different mass density are connected, wave pulses on one string are *both* partially transmitted onto the other string *and* are partially reflected from the boundary. Computing the transmitted and reflected waves is straightforward but beyond the scope of this class (it starts to involve real math and studies of boundary conditions). However, the following qualitative properties of the transmitted and reflected waves should be learned:

- Light string to heavy string (most "like" a fixed point): Transmitted pulse right side up, reflected pulse inverted. (A fixed string boundary is the limit of attaching to an "infinitely heavy string").

- Heavy string to light string (most "like" a free point): Transmitted pulse right side up, reflected pulse right side up.

It is often difficult to visualize what traveling waves, including wave pulses, on a string look like because it is very difficult to build a taut string that stretches indefinitely to the right or left and put a traveling wave onto it. One has to connect the string to *something* at both ends and *pull it tight!* to get it to where our analysis above makes sense!, and as we've seen, this imposes boundary conditions where it is tied, or held, or shaken

However, it is easy enough to program a computer to literally *solve Newton's second law* for mass "chunks" connected by "springs" in a linear "string" to illustrate traveling wave pulses and more. Some clever physics/computer geeks have done so, and created a lovely "virtual lab" accessible on the internet for students to freely use to explore their properties!

I encourage all of the students reading this text to visit at least the following online "lab":

https://phet.colorado.edu/sims/html/wave-on-a-string/latest/wave-on-a-string_en.html

to take a look and spend some time "playing" with its very intuitive interface. Initially, just generate wave pulses and watch them propagate, but be sure to go back and simulate *all* of the solutions we develop in this chapter, especially the ones that come next: Harmonic waves, both traveling and stationary!

10.5: Harmonic Traveling Waves

So far, the waves we've studied don't even *look* like waves, at least not the waves we tend to *call* waves. Waves at the ocean, for instance, come in one after the other, sort-of periodically, rising and falling with some distance between the crests. If you have learned much about music, you might know that middle C on the standard scale is a sound wave that has a frequency of 256 hertz (cycles per second). You might have looked at a radio tuner and noted that the frequencies of radio waves transmitting music or other common signals are anywhere from 500 kilohertz to hundreds of megahertz. You've probably used terms like frequency and wavelength before, and wave pulses don't *seem* to have any of that[254]

Enter ***harmonic traveling waves***; waves that *look* like waves and have things like frequency and wavelength!

The harmonic functions are (in case you didn't know this already) our old friends ***sine*** and ***cosine***. Harmonic waves are therefore simply (obviously twice differentiable) waveforms where:

$$f(u) = A\sin u \quad \text{or} \quad f(u) = A\cos u$$

Simple, right?

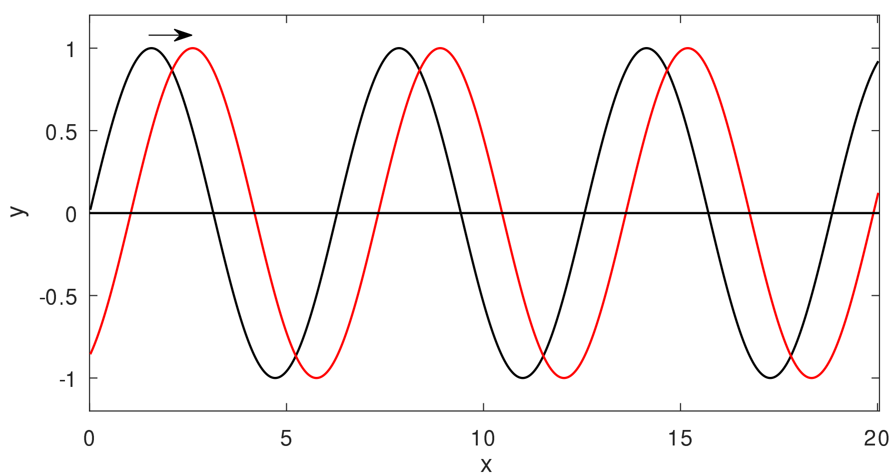


Figure 10.5: A *harmonic traveling wave* "snapshot" at $vt = 0$ (in black, plus a second snapshot of the wave 1/6th of a temporal period later.

Now, however, we have to use these *functions* to *solve the 1DWE*, so we have to use $u = x \pm vt$ as their argument. For example:

$$y(x,t) = y_0 \sin(x - vt) \tag{10.27}$$

is one such wave. In figure 10.5, I've plotted this very wave form at two distinct times: $vt = 0$ (in black) and $vt = \pi/3$ (in red). Note how the entire waveform has shifted to the right by 1/6th of a 2π cycle, as expected!

[254]They do, but you have to learn Fourier analysis to learn how, and this chapter is a first step in that direction...

In this figure, I used $y_0 = 1$ (in whatever length unit you like), say centimeters) because octave requires an actual scale to work with. In general, y_0 is a constant set from information given with the problem (or a placeholder for that information) and is called the *amplitude* of the harmonic wave. As the name suggests, it plays the exact same role as the amplitude of a harmonic oscillator (which you hopefully understand reasonably well at this point), so we won't say much about it here beyond that. But what about the wave itself? There doesn't seem to be any *choice* that would let us set wavelength or frequency or phase as we expect to be able to do!

One thing is immediately clear: the solution we graphed above ***is missing a scale factor*** needed to ***make the argument of the sine (or cosine) function dimensionless*** as discussed in a previous section.

The *purpose* of this scale factor for harmonic solutions is to convert u in units of length (along the string, recall) into dimensionless *radians*, the angular argument of sine or cosine (or other trig functions). Both harmonic functions have an *angular* period of 2π. This motivates the definition of a *spatial* period for a harmonic wave, called ***the wavelength***, to use in the required conversion factor.

We will use the symbol λ to represent wavelength from here onward – this is a very conventional choice made in many textbooks[255] with implicit units of length, e.g. meters in the SI system. Note well that many *different* units may be used for wavelengths in different contexts, as most waves of interest to anyone but a surfer will have wavelengths *much smaller* than a meter.

To convert wavelengths into radians, then, we need to multiply $u = x - vt$ by 2π divided by one wavelength. We give this quantity its *own* special name and symbol. It is called the (angular) ***wave number***[256] and its standard symbol is k:

$$k = \frac{2\pi}{\lambda}$$

and it does, indeed, have units of one over length (radians per meter, but radians themselves are dimensionless).

Now we are in a better position to write a *completely general* harmonic solution to the 1DWE:

$$y(x,t) = y_0 \sin(k(x \pm vt) + \phi) = y_0 \sin(kx \pm kvt + \phi)$$

Note that I didn't box this – yet – because there are a couple more tweaks to make to it first.

The first is to answer the question: "But what about cosine? How is this general?" The answer should be familiar to you from the last chapter – the phase angle ϕ is a *constant of integration* (as is the wave amplitude y_0) just as it was for harmonic oscillation. Indeed, we could have used the exponential solution assumption we used in the last chapter and solved the 1DWE *directly* for harmonic traveling waves as the real part of complex exponential forms (and will do something similar to this shortly when we discuss standing waves). By adjusting ϕ to e.g. $\pm\pi/2$, we can convert sine into cosine or vice versa, so we don't really need both.

[255]Wikipedia: http://www.wikipedia.org/wiki/Wavelength. As well as wikipedia...

[256]Wikipedia: http://www.wikipedia.org/wiki/wavenumber. Most physicists skip the "angular" bit and just call k the wave number, but it does make the usage a bit more consistent to include it and emphasize the connection to the *angular* (temporal) *frequency* ω we are already familiar with.

The second is to recall that *from* our work in the last chapter, we *know, in principle, how to evaluate* ϕ should we ever need to. Or to invert this process, we can *always* choose to set our reference clock such that $\phi = 0$ for either sine or cosine solutions, so we might as well just choose one function or the other and not bother to include the phase angle ϕ at all, knowing we can always stick one in if needed.

The third is to note that our solution oscillates in space, to be sure, at any fixed time, but it also ***oscillates just like a harmonic oscillator*** at any *fixed* point in space! In that case 'kx' acts like a fixed phase and the solution would look like e.g. $y_0 \sin(-kvt + \text{`}kx\text{'}) \sim -y_0 \sin(\omega t + \phi)$ where we have identified kv as ω, the ***angular frequency of oscillation*** in the harmonic wave.

Putting it all together, for a general harmonic wave traveling to the **right**:

$$\begin{aligned} y(x,t) &= y_0 \sin(k(x - vt) + \cancel{\phi}) && (10.28) \\ &= y_0 \sin(kx - kvt) && (10.29) \\ &= y_0 \sin(kx - \omega t) && (10.30) \end{aligned}$$

Of course we can equally simply construct the harmonic wave traveling to the left. Our general *enough* harmonic traveling wave solution(s) are then:

$$\boxed{y(x,t) = y_0 \sin(kx \mp \omega t)} \qquad (10.31)$$

where the minus sign means the wave train propagates to the right, the plus sign the left, and we *can* add a phase back to either one but usually won't have any need to. And we're done!

In these solutions:

$$kv = \frac{2\pi}{\lambda} v = \omega = 2\pi f = \frac{2\pi}{T} \qquad (10.32)$$

(the last two relations inherited from the last chapter) and where (with a bit of algebraic manipulation and rearrangement) we see that we have *several* ways to write v:

$$v = f\lambda = \frac{\omega}{k} = \frac{\lambda}{T} \quad \left(= \sqrt{\frac{\mathcal{T}}{\mu}} \right) \qquad (10.33)$$

where f is the temporal frequency in Hertz (inverse time, cycles per second) and T is the temporal period of the wave, the temporal equivalent of the spatial period λ. I connected everything back to v determined from the 1DWE itself for this specific case of waves on a string. This is the *only* thing that isn't quite general to *all harmonic waves* in any context in physics.

Note that I used a different font here for tension $\mathcal{T}$ and the ***temporal period of the wave*** T, as unfortunately capital 'T' is typically used for both, often in a single problem. With practice you will keep them straight without thinking about it, but when learning try to differentiate them in some way so you don't e.g. try to cancel $\mathcal{T}$ with T!

Note Well: You should simply ***know every relation*** in this set of algebraic relations between $\lambda, k, f, T, \omega, v, \mathcal{T}, \mu$ to save time on tests and quizzes. Yes, you only *need* to know a *very few* of the relations and definitions and you can *derive* or *fill in* the rest. Normally, I'd suggest only learning this minimal set (but learning it *very well*)and then rely on your ability to quickly derive all of the rest of it on the fly, but for waves *I do not* because it takes *time*, time you'd rather be putting them to work on a quiz or exam. Besides, they are the basic vocabulary of waves in all contexts.

To make learning them simple, I've collected "what you need to just know" about harmonic waves on a string into a ***single box***:

$$\boxed{\begin{array}{c} k = \dfrac{2\pi}{\mathcal{T}} \qquad k = \dfrac{2\pi}{\lambda} \qquad kv = \omega \qquad f = \dfrac{1}{\mathcal{T}} \\ v = \dfrac{\omega}{k} = \dfrac{2\pi/\mathcal{T}}{2\pi/\lambda} = \dfrac{\lambda}{\mathcal{T}} = f\lambda = \sqrt{\dfrac{T}{\mu}} \end{array}}$$

Harmonic traveling waves are extremely important in nature, medicine, and human affairs – not waves on a string, particularly, but harmonic *sound* waves – music! –, harmonic *light* waves (including e.g. radio waves), harmonic *quantum wavefunctions*, harmonics in *Fourier analysis*. Harmonic traveling waves on a string are merely our *first look* at harmonic waves, but future physicist or future physician alike, you will almost certainly need to understand "generic" wave theory well enough to be able to understand how e.g. ultrasound or radar work, how wave diffraction limits resolution of visible light microscopes, how radios work, and so on.

Let's leave traveling waves for the moment – we'll come back to them shortly in a further discussion of wave pulses – and look at a second *kind* of harmonic solution to the wave equation, one of extreme importance in the construction of e.g. stringed musical instruments and one that exhibits *resonance*: Stationary harmonic waves.

10.6: Stationary Harmonic Waves

The third special case of solutions to the 1DWE that we will cover is that of ***stationary harmonic waves***, also called ***harmonic standing waves***. Note well "stationary" doesn't mean that the string is frozen in the shape of a wave – it moves! Stationary harmonic waves are stationary in the sense that the ***oscillating harmonic waveform does not move to the left or the right down the string!***

Stationary waves arise when we stretch some given string of finite length L connecting it one of two distinct ways to a rigid object at *both* ends that can keep it under tension. These ways are identical to those we considered in the section on wave pulse reflection. First, fastening the string to a fixed hook connected to a wall, to a peg, to a nail so that the string at the connection point ***does not move*** as the string oscillates in waves. Second, fastening the string to a nearly massless, nearly frictionless ring and threading the ring onto a nearly frictionless rod held rigidly at right angles to the string in such a way that the normal force between rod and ring keeps the string under tension.

We refer to these as "fixed" or "free" ends of the string, respectively. We must then develop solutions for the finite string of length L that is kept stretched at the tension $\mathcal{T}$ by being fixed *or* free at *both* ends, or else is kept stretched by being fixed at one end and free at the other (four distinct possibilities). The constraints on value or slope at the fixed or free ends are called the ***spatial boundary conditions*** on the wave solution to the 1DWE, and they result in standing wave solutions.

In a moment, we'll *formally* derive standing wave solutions (it isn't too difficult if you remember SHOE from the last chapter) but first, let's think a moment see if we can figure out what they must look like a different way.

Suppose we have a string of length L that is fixed at both ends – think ‘a guitar string’ (or harp, or piano, or violin or fill in your favorite stringed instrument here). If we put a harmonic wave travelling to the right on the string, it has to hit the fixed end, and it has no choice but to reflect. This is because the *energy* in the waveform on the string cannot just disappear, and if the end point is fixed no work can be done by it on the peg to which it is attached and that keeps the string taut. The reflected wave has to have the same amplitude and frequency as the incoming wave and travels to the left, where it hits *that* peg and reflects.

It seems, then, as though any wave that can survive on a string fixed at both ends must be a superposition of *two traveling waves* – one going to the right and one going to the left – with identical frequency, wavelength, and amplitude. This superposition will still solve the 1DWE because of the superposition principle, but we have no idea (yet) what that superposition would look like. In particular, we need to see if we can make the function that results from adding the two waves satisfy the *particular* boundary conditions at the ends of the string:

$$y(0,t) = 0 \qquad y(L,t) = 0$$

that describe the string connected to immovable pegs at the ends. These are called *Dirichlet* boundary conditions in the general field of boundary value problems.

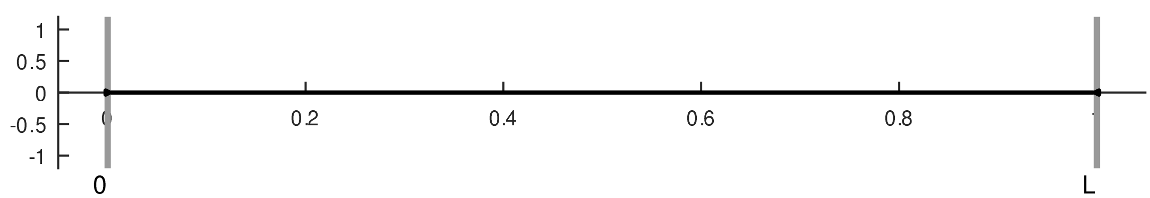


Figure 10.6: A “blank” stretched string, fixed at both ends, running from 0 to L on the x axis. There is no wave on it yet (or if you prefer, there is a wave with amplitude 0, as this *does* satisfy the 1DWE but is *boring!*

Figure 10.6 displays a simple “blank” graph that we will use to explore and understand standing waves once we find the ones that satisfy the required boundary conditions on a string. This one is ***fixed at both ends***, as that is the first, and simplest, case and the one you almost certainly have encountered many times in your life already.

Now we need to add two waves with the same amplitude, opposite directions, and the same frequency and wavelength as suggested by our *intuitive* discussion above. We won’t worry about whether the two waves are relatively inverted – it won’t matter as we’ll see later that we are really getting a *family* of solutions and will have freedom to choose the one we need to make solving the boundary conditions as simple as possible. However we do have to remember how to add two distinct sine functions! I therefore remind you of a trig identity you probably learned long ago[257]:

$$\sin A + \sin B = 2 \sin\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right)$$

Then (using y_0' as the amplitude of the two waves):

$$\begin{aligned} y(x,t) &= y_0' \sin(kx - \omega t) + y_0' \sin(kx + \omega t) && (10.34)\\ &= y_0' \left(\sin(kx - \omega t) + \sin(kx + \omega t)\right) && (10.35)\\ &= 2y_0' \sin(kx)\cos(\omega t) && (10.36)\\ &= y_0 \sin(kx)\cos(\omega t) && (10.37)\end{aligned}$$

where we rename $2y_0'$ as the *final* amplitude of the resulting wave form y_0 for consistency's sake. This is perfectly all right, since we will be setting y_0 directly from the givens and will never actually "see" the amplitudes y_0' of the left and right propagating waves.

Note well this wave form ***does not*** travel to the right or the left on the string. Each bit of string at position x oscillates harmonically at angular frequency ω through the amplitude determined by $y_0 \sin kx$! At some values of x, $\sin(kx) = 0$ *at all times* and the string is *indeed* stationary – these fixed points are called *nodes*. At other values of x, the string oscillates through the full value of A – these are called *antinodes*. Our job will be to align this solution with the x axis in such a way as to guarantee nodes where we want them and antinodes where we want *them* (they're associated with *free* boundary conditions, also called *Neumann* conditions). It will turn out momentarily that this is possible only for an infinite list of discrete frequencies/wavelength.

Recall that the traveling wave solutions we added above can independently contain arbitrary phases that we could use to turn either of the sine functions into a cosine function (still ensuring that we add two waves going in opposite directions). When we apply the addition theorem, we could put these "extra" phases whereeever we like, *or* use *other* forgotten trig identities to show that the particular solution we got above is just a special case of a more general family:

$$y(x,t) = y_0 \cos(kx + \delta)\cos(\omega t + \phi) \tag{10.38}$$

where by suitable (free) choices for ϕ and δ we can turn the cosine into sine or produce a minus sign, just as we could for the SHOE solutions from the last chapter to accommodate different initial (not boundary!) conditions. In the next section, we will formally derive this solution without all of the handwaving and use of heuristic arguments above.

As usual, physics majors and math majors and engineers should go over this section in complete detail; life science students are *advised* to walk through it (I will, in class) but will not be responsible for learning it or reproducing it. Frankly, physics majors won't either – yet – but the day will come when they must so starting to learn it now will help a lot on that day.

[257] In my case, *very* long ago. This result is actually pretty easy to derive using by adding two special vectors called "phasors" – no relation to Star Trek weaponry – a methodology that works for *more* cases than are handled by the trig identity alone. I'll show you how this works later in the course.

10.6.1: Formal Derivation of Standing Wave Solutions: Separation of Variables

So far, we solved the 1DWE for traveling wave solutions by introducing a *single* variable containing *both* x and t (and showing that pretty nearly any "nice" function was then a solution). But could there not be solutions where x and t appear *independently?* If our heuristic reasoning above did nothing else, it showed that the answer to this, at least, is yes.

Inspired by the fact that a standing wave solution with imposed Dirichlet[258] or Neumann[259] is ***necessarily*** a product of a spatial part that can always have a specified value at some point and a temporal part that (we can at least guess) oscillates more or less harmonically we can *derive* solutions that have this form by assuming that the function $y(x,t)$ can be written as the *product* of a function of x alone and a second function of t alone in the 1DWE, and then solving for those functions. This approach – called *separation of variables* – to solving differential equations is quite general, and while it doesn't *always* work often it does in cases of interest in physics, mathematics, and engineering. We've actually used it implicitly several times already for simpler differential equations e.g. those that describe linear drag, but in those cases we could simply integrate a first order equation.

We therefore assume that:

$$y(x,t) = \mathrm{X}(x)\mathrm{T}(t) \tag{10.39}$$

Note that I made both X and T with a *times roman* font to differentiate it from *both* period T *and* string tension $\mathcal{T}$. Sigh.

If we substitute this into the differential equation, use the fact that time derivatives only act on $\mathrm{T}(t)$ and space derivaties only act on $\mathrm{X}(x)$, and divide by $y(x,t) = \mathrm{X}(x)\mathrm{T}(t)$ on both sides to get all the t dependent parts on the left and all of the x dependent parts on the right, we get:

$$\frac{d^2y}{dt^2} = \mathrm{X}(x)\frac{d^2\mathrm{T}}{dt^2} = v^2\frac{d^2y}{dx^2} = v^2\mathrm{T}(t)\frac{d^2\mathrm{X}}{dx^2} \tag{10.40}$$

$$\frac{1}{\mathrm{T}(t)}\frac{d^2\mathrm{T}}{dt^2} = v^2\frac{1}{\mathrm{X}(x)}\frac{d^2\mathrm{X}}{dx^2} \tag{10.41}$$

$$= -\omega^2 \tag{10.42}$$

where the last line follows because in the second line, a term containing t *only* is set equal to a term containing x *only*. These can be independently varied in our solution, so ***both terms must equal a constant!***

Why did I choose a *negative* constant, $-\omega^2$? Because – gifted with second sight (or if you prefer, armed by experience) – I can see four lines of algebra into the future and know that this will give us the family of solutions we need, but *I could have chosen it to be positive* and we could still solve the resulting separated differential equations. However, in this case the solutions would look quite different (like real exponential functions, in fact) and *would not work* to describe wave oscillations on a string subject to the required boundary conditions! We encountered something similar last chapter and I'll say no more about this, but in a future differential equations course you might encounter this alternative.

[258]Wikipedia: http://www.wikipedia.org/wiki/Dirichlet boundary condition. Interested students can pursue this further here...
[259]Wikipedia: http://www.wikipedia.org/wiki/Neumann boundary condition. ...and here.

For *us, now*, then, we have turned our second order differential equation in *two* variables into *two* differential equations, each with a *single* (but different) variable! Hence "separation of variables"! We get:

$$\frac{d^2\mathrm{T}}{dt^2} + \omega^2\mathrm{T} = 0 \tag{10.43}$$

and

$$\frac{d^2\mathrm{X}}{dt^2} + k^2\mathrm{X} = 0 \tag{10.44}$$

(where we use $k = \omega/v$). Do you recognize them? I'm certainly *hoping* you do without further prompting!

However, this was a rhetorical question in context, so I'll go ahead and answer it.***These are both SHOEs*** (in appropriate coordinates) so we don't actually *have* to solve them, we already *know* their solutions (from the last chapter)! Indeed, we already know their solutions from what we wrote down heuristically up above!

They are:

$$\mathrm{T}(t) = \mathrm{T}_0\cos(\omega t + \phi) \tag{10.45}$$

and

$$\mathrm{X}(x) = \mathrm{X}_0\cos(kx + \delta) \tag{10.46}$$

so that the most general stationary solution can be written:

$$y(x,t) = y_0\cos(kx + \delta)\cos(\omega t + \phi) \tag{10.47}$$

exactly as we already wrote it!

The x dependent part of this, including the amplitude y_0, we will often refer to as the *envelope* of the standing wave – the maximum extent of the string from its straight position at each point. We'll usually draw this along with its negative to help us visualize the oscillation.

We will in this class *always* use δ phases such that the amplitude function is either a pure sine or pure cosine function (omitting δ altogether either way) as they will make it maximally simple to match our solutions to Dirichlet or Neumann boundary conditions imposed at $x = 0$. For the time dependent part we will always by convention set $\phi = 0$ and just use $\cos(\omega t)$ as the time dependence. As noted, this is just like setting our reference clock to be zero just as the string comes momentarily to rest at its maximum envelope position. Hence our solutions will always look like either:

$$y_0\sin(kx)\cos(\omega t) \quad \text{or} \quad y_0\cos(kx)\cos(\omega t)$$

The choice of which one we should use for each kind of problem use will be illustrated shortly in context.

A final comment for the mathematically inclined: Imposing boundary conditions on solutions to this second order differential equation caused them to become *quantized* – only occur for an (often infinite) set of discrete values for e.g. the frequency and/or the wavelength. This happens under quite general circumstances, and you can pursue the idea further starting at:

Wikipedia: http://www.wikipedia.org/wiki/Sturm-Liouville theory (or in your favorite differential equations course). It is the basis of discrete energy levels and more in quantum mechanics. Once again this may just be your *first* glimpse into this but not your last!

We now return to your regularly scheduled coverage carrying just one idea that we already had in hand but have now formally derived – harmonic stationary wave solutions are built out of two harmonic functions, one of kx only and one of ωt only. This is enough.

10.6.2: Standing Wave Solutions: Fixed Boundary Conditions At Both Ends

There are *four* distinct patterns of boundary condition possible for our string. We can have:

$$y(0,t) = 0 \qquad \text{or} \qquad \left.\frac{dy}{dx}\right|_{x=0} = 0 \tag{10.48}$$

$$y(L,t) = 0 \qquad \text{or} \qquad \left.\frac{dy}{dx}\right|_{x=L} = 0 \tag{10.49}$$

Any of the four permutations of these conditions at 0 and L are permitted. We'll look at them all. We wish to be able to take any "spanning set" of information about the string – some combination of v, $\mathcal{T}$, μ, L, amplitude y_0 and perhaps information to identify a particular *mode of oscillation* and fill in a table with – everything *else* there is to know about the oscillating string. Frequency, wavelength, angular frequency, wave number – stuff from the box above. Keep it close at hand as you work through this chapter.

Before proceeding, let me note that for stringed instruments or other resonant harmonic systems, the lowest frequency, produced by the longest possible wavelength, is called the ***principle harmonic*** (or ***fundamental frequency***) and is the frequency that normally *gives the string its pitch – the "note" it plays when plucked.* However, plucking the string puts energy in many of the other ***modes*** permitted by the string's boundary conditions as well! The particular superposition of these ***harmonics*** are what give guitar strings, harp strings, piano strings their particular tone, and reproducing this *mix* of tones is what electric keyboards do with switches and internal tone generators that can make them sound like a piano, an organ, a guitar, or (my own favorite) screaming valkyrie angels when a key is pressed!

In the discussion below I will use two terms that are easily, and often, confused*even in physics textbooks or problem solutions* – mode and harmonic. The (resonant!) ***modes of oscillation*** are all of the permitted harmonic waveforms that satisfy the boundary conditions. These are true resonances – if the string is *driven* with a harmonic driving force or wave at one of the mode frequencies, that mode gains energy like a driven harmonic oscillator in resonance. We will often refer to modes counted in order from the lowest frequency, longest wavelength fundamental mode through each successively higher resonant frequency.

Harmonic refers explicitly to the ***nth integer multiple of the fundmental frequency***. If the fundamental frequency or principal harmonic is f_1, he third harmonic frequency is thus $3f_1$. For certain boundary condition combinations (like fixed at both ends) the two are basically synonymous – the 3rd harmonic is also the frequency of the 3rd mode. However, for others they are *not*, and we will need to identfy the difference.

We'll start with the example of a string fixed at both ends, as we're more than halfway to the answers for it already. The boundary conditions in this case are:

$$y(0,t) = 0 \qquad \text{and} \qquad y(L,t) = 0$$

We note that if we choose $\sin(kx)$ as our spatial function, we are *guaranteed* to satisfy the left hand boundary condition at $x = 0$, so we'll do that. The form of our solution is therefore:

$$y(x,t) = y_0 \sin(kx)\cos(\omega t)$$

The sine function forms an *envelope* such that each little chunk of the string oscillates harmonically between $\pm y_0 \sin(kx)$ limits. In order to end up with the required node at $x = L$, we need:

$$y(L,t) = y_0 \sin(kL)\cos(\omega t) = 0 \tag{10.50}$$

at all times, meaning that the spatial sine function has to be zero when $x = L$!

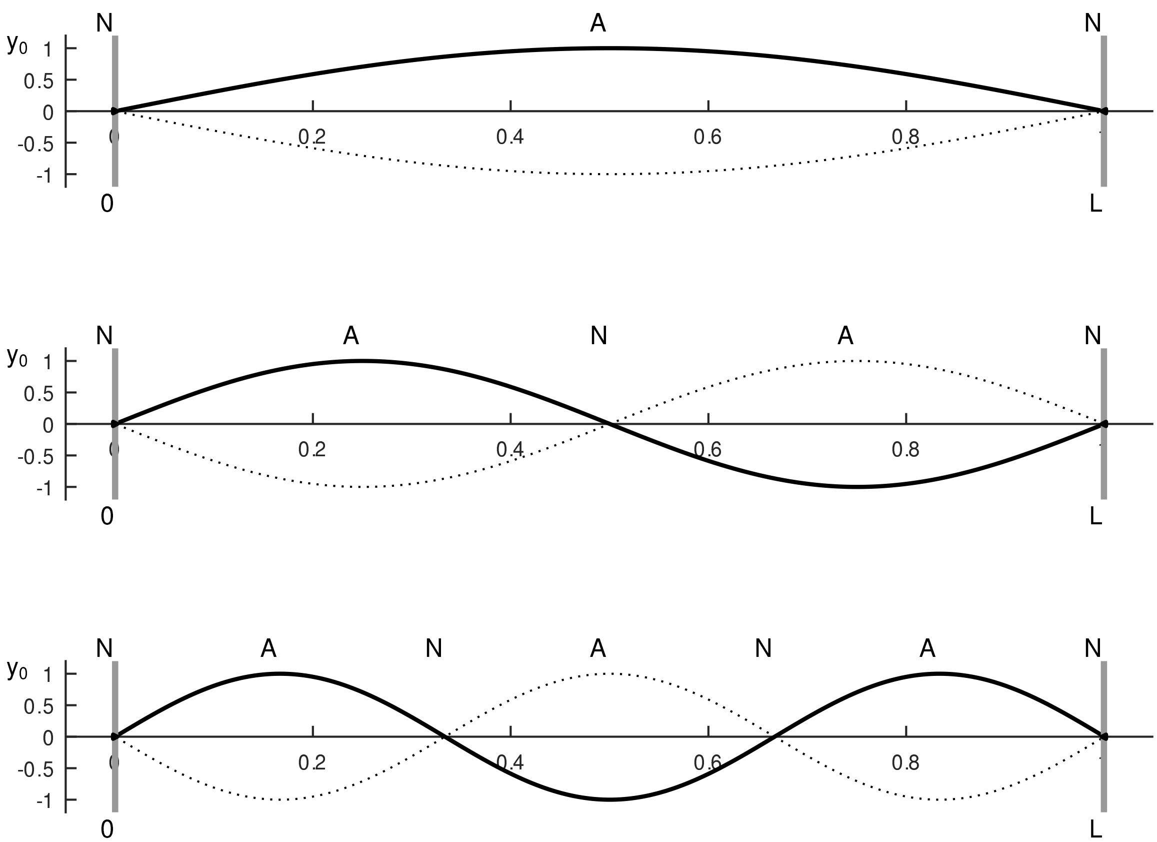


Figure 10.7: The first three modes of a string fixed at both ends, with nodes (N) and antinodes (A) labelled. Note well the nodes at the ends!

Hopefully you recall that:

$$\sin\theta = 0 \quad \text{when} \quad \theta = 0, \pi, 2\pi, 3\pi..., = n\pi... \quad \text{for} \quad n = 1, 2, 3...$$

Apparently can therefore satisfy our boundary condition at $x = L$ only for certain *discrete* values of k (or λ). We can find them from the equation(s):

$$k_n L = \frac{2\pi L}{\lambda_n} = \pi, 2\pi, 3\pi, \ldots = n\pi \text{ for } n = 1, 2, 3... \tag{10.51}$$

We skipped $n = 0$ because that leads to *no wave on the string*. This is the *trivial* solution to the 1DWE – correct but useless! Some simple rearrangment allows us to solve for λ_n:

$$\lambda_n = \frac{2L}{n} \tag{10.52}$$

for $n = 1, 2, 3, ...$ The first three modes in the *countably infinite* series of modes are portrayed in figure 10.7 above. This result is true *independent* of the mass density of the string or its tension.

Once you see the pattern in the figure above, you can basically ***predict the allowed wavelengths by drawing them onto a string***. To me, at least, the oscillation envelope for the nth harmonic looks like a string of n link "sausages" packed into the distance from 0 to L. Use this mental "image to draw the 5th harmonic *without using any formula or math!* Visualize five juicy sinusoidal sausages in row, stretched along the axes and draw them there. To get the link distances right, divide L up into five equal chunks of length $L/5$ each. Label the middle of each sausage with an 'A' (there should be 5 A's). Label the links with an 'N' (there should be six N's, with one at both ends). And you're done! You can now *look at your figure and see the wavelength!* It is the length of two sausages! Your graph should look like a hand-drawn version of figure 10.8.

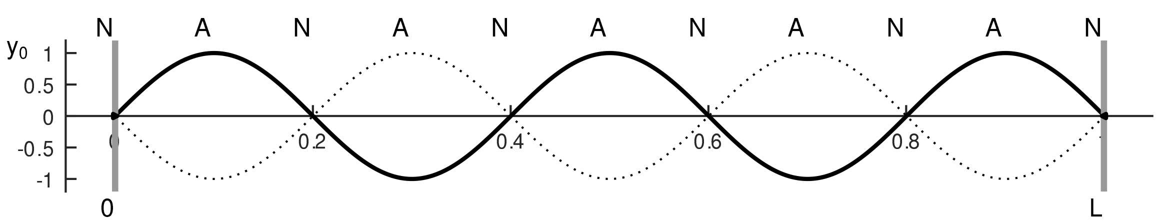


Figure 10.8: Five half-wavelength "sausages" arranged along 0 to L. The wavelength of this ***fifth harmonic*** is (hopefully obviously) $\lambda_5 = \frac{2}{5}L$.

With a formula for the wavelength of the nth mode (useful) in hand and a sausage-y mnemonic in hand powerful enough that you may not even *need* the formula on a quiz or exam, we can proceed to answer the question. Let's obtain the harmonic frequencies of the first three modes. We'll use the ***formulas from the box*** which I will from now on use freely without further discussion, assuming you know them!

$$f_1 = \frac{v}{\lambda_1} = \frac{v}{2L} = \boxed{\frac{1}{2L}\sqrt{\frac{\mathcal{T}}{\mu}}} \tag{10.53}$$

$$f_2 = \frac{v}{\lambda_2} = \frac{2v}{2L} = \boxed{\frac{1}{L}\sqrt{\frac{\mathcal{T}}{\mu}} = 2f_1} \tag{10.54}$$

$$f_3 = \frac{v}{\lambda_3} = \frac{3v}{2L} = \boxed{\frac{3}{2L}\sqrt{\frac{\mathcal{T}}{\mu}} = 3f_1} \tag{10.55}$$

...

$$f_5 = \frac{v}{\lambda_5} = \frac{5v}{2L} = \boxed{\frac{5}{2L}\sqrt{\frac{\mathcal{T}}{\mu}} = 5f_1} \tag{10.56}$$

...

I went ahead and filled in f_5 since we already got its wavelength. This is easy! The frequency of the nth *mode* is n times the frequency of the *first* mode! We could have just done:

$$f_n = \frac{v}{\lambda_n} = \frac{nv}{2L} = n\frac{1}{2L}\sqrt{\frac{\mathcal{T}}{\mu}} = nf_1$$

once and for all at the very beginning! Hopefully you see how easy it is now to find, for example, $\omega_n = 2\pi f_n$, $k_n = 2\pi/\lambda_n = n\pi/L$, $T_n = 1/f_n$, etc. The box rules!

Example 10.6.1: A String Fixed at Both Ends

Let's do an example. Suppose we are given a string with mass density $\mu = 0.00033$ kg/m, length $L = 1$ m, that is fixed at both ends. This might well physically represent a guitar string, where the string is "fixed" in slots at the "saddle" and "nut" at the ends of the neck and body. We are asked to find the tension $\mathcal{T}$ we need to put on the string to tune it to 1024 Hz (high E), and then describe the waves associated with the next two ***harmonics*** associated with the string. As always, we'll solve this problem algebraically first, and only then substitute the numbers.

Most of the *solution* is given above. Once one has $\lambda_1 = 2L$ (from the formula or by drawing a picture), one finds f_1 in terms of the givens and solves for $\mathcal{T}$. Since we *just did this* I'll go ahead and start with the formula for $f_1 = v/\lambda_1$, do a tiny bit of algebra:

$$f_1 = \frac{1}{2L}\sqrt{\frac{\mathcal{T}}{\mu}} \quad \Rightarrow \quad \mathcal{T} = 4L^2 f_1^2 \mu$$

and then plug in the given numbers:

$$\boxed{\mathcal{T} = 4 \times 1^2 \times 1024^2 \times 0.00033 = 1397 \text{ newtons}} \tag{10.57}$$

Note that a) these numbers are realistic; and b) this is a considerable tension – 312 pounds in barleycorn units. It's a lot greater than *my* weight, and I'm 6'1' and not particularly skinny[260]!

[260] A polite way of saying that I'm, um, chubby? Obese? Pudgy? All of the above?

The neck of a guitar has to be quite strong to withstand the stress, and the string density has to vary across its *width* as well to ensure that the tension of the *low* E string at least approximately equals that if the high E string, or the imbalanced tension exerts a *sideways* torque that, over time, can bend the neck! Most guitar necks contain a "truss rod", a screw-adjustable *metal rod* that runs the length of the neck and can be (slightly) expanded or contracted with a wrench to counteract upward or downward bowing of the neck under compressive stress equivalent to the weight of a medium size grizzly bear or small cow.

10.6.3: Standing Wave Solutions: Free Boundary Conditions At Both Ends

It is also[261] possible to stretch a string so that it is fixed at one end but so that the *other* end is *free to move vertically*. In this case, instead of having a node at the free end (where the wave itself vanishes), the *slope* of the wave at the end has to vanish. We need this to be true at *both* ends:

$$\left.\frac{dy}{dx}\right|_{x=0} = 0 \qquad \left.\frac{dy}{dx}\right|_{x=L} = 0 \tag{10.58}$$

Basically, this just means that our solution has to have ***antinodes at both ends***

The primary change in what we did above is that we now need to use *cosine* instead of *sine* as the spatial wave form. Since $\cos(0) = 1$, that puts an ***antinode at both ends!***

There are then two ways to get to the sequence of λ_n for this set of boundary conditions. The first is easy and intuitive "sausagespeak" – need to pack links into the 0 to L distance so that our string of links *begins and ends with a half sausage* and so that there is *at least one node/link in between*. With no more of a rule than that, you *should* be able to draw the first three modes for the string starting at the fundamental mode/principle harmonic, determine the wavelength(s) and literally fill in everything else from some set of givens using the formulas in the box.

The second is to do the math. Using:

$$y(x,t) = y_0 \cos(kx)\cos(\omega t) \tag{10.59}$$

guarantees that there is an antinode at $x = 0$. How do we get a node at the other end? Recall that

$$\cos(\theta) = \pm 1 \quad \text{when} \quad \theta = 0, \pi, 2\pi, 3\pi ... \quad \text{or} \quad \theta = \theta_n = n\pi \quad \text{for} \quad n = 1, 2, 3...$$

We have to use this and:

$$y(L,t) = \pm y_0 = y_0 \cos(k_n L)\cos(\omega_n t) \tag{10.60}$$

as before:

$$k_n L = \frac{2\pi}{\lambda_n} L = n\pi \qquad \Rightarrow \qquad \lambda_n = \frac{2L}{n} \quad \text{with} \quad n = 1, 2, 3... \tag{10.61}$$

(skipping $n = 0$ as before as it leads to the trivial solution).

This is *exactly the same sequence as a string fixed at both ends!* All we have to do is graph out the sequence of "sausages" in figure 10.9, read off λ_n, and use the box to answer

[261] In principle, not so much in practice...

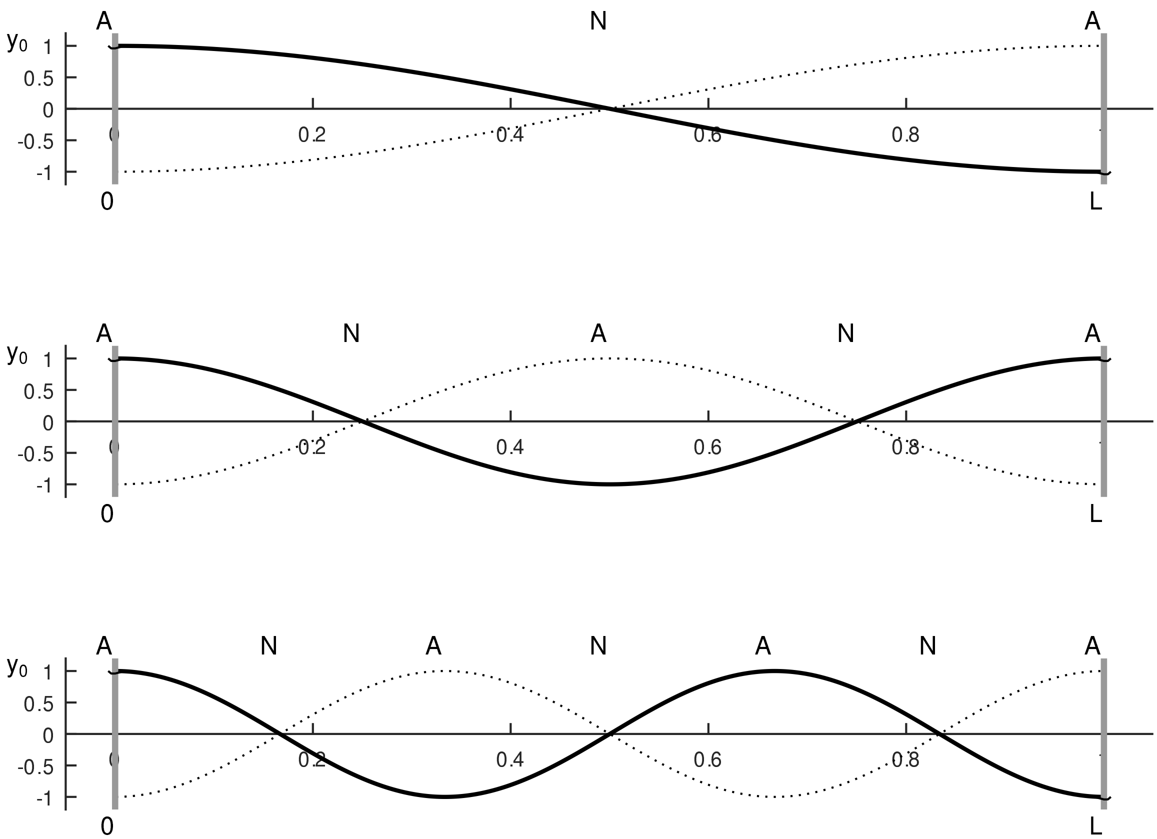


Figure 10.9: The first three modes of a string free at both ends, with nodes (N) and antinodes (A) labelled. Note well the antinodes at the ends!

any questions in an actual problem, without even *having* to remember:

$$\lambda_n = \frac{2L}{n} \qquad n = 1, 2, 3...$$

I'm not going to recapitulate the math for f_1, f_2, f_3... etc. It is identical to that in the previous section.

Note well: The mode index is identical to the harmonic index – the third mode is also the third harmonic, just as it was for fixed and fixed boundary conditions. ***That will not be the case for the next two possible boundary conditions!***

10.6.4: Standing Wave Solutions: Fixed and Free Boundary Conditions At Opposite Ends

I'm going to do both of these at once. By now you should get the idea that a Dirichlet end has a node and requires a "sausage" link, while a Neumann end has an antinode and requires a terminating half-sausage. With no more information than that, try to draw in the first three

modes, for boundary conditions of:

$$y(0,t) = 0 \qquad \left.\frac{dy}{dx}\right|_{x=L} = 0 \tag{10.62}$$

You should get a half sausage, with the cut at $x = L$ for the first one. Since a half sausage is a quarter of a wavelength, $\lambda_1 = 4L$! If we use the other order, free and fixed:

$$\left.\frac{dy}{dx}\right|_{x=0} = 0 \qquad y(L,t) = 0 \tag{10.63}$$

you should *still* have a half sausage, but now the cut end is at $x = 0$ and the link at $x = L$. The wavelength is *still* $\lambda_1 = 4L$.

For the second mode supported by the string, we need one more node and antinode. But that only adds a single sausage, to put one and a half sausages on the string, link at $x = 0$, half sausage at $x = L$. The third mode adds one more – link, two full sausages, and a half at the end. Can you see what the wavelengths must be for these two?

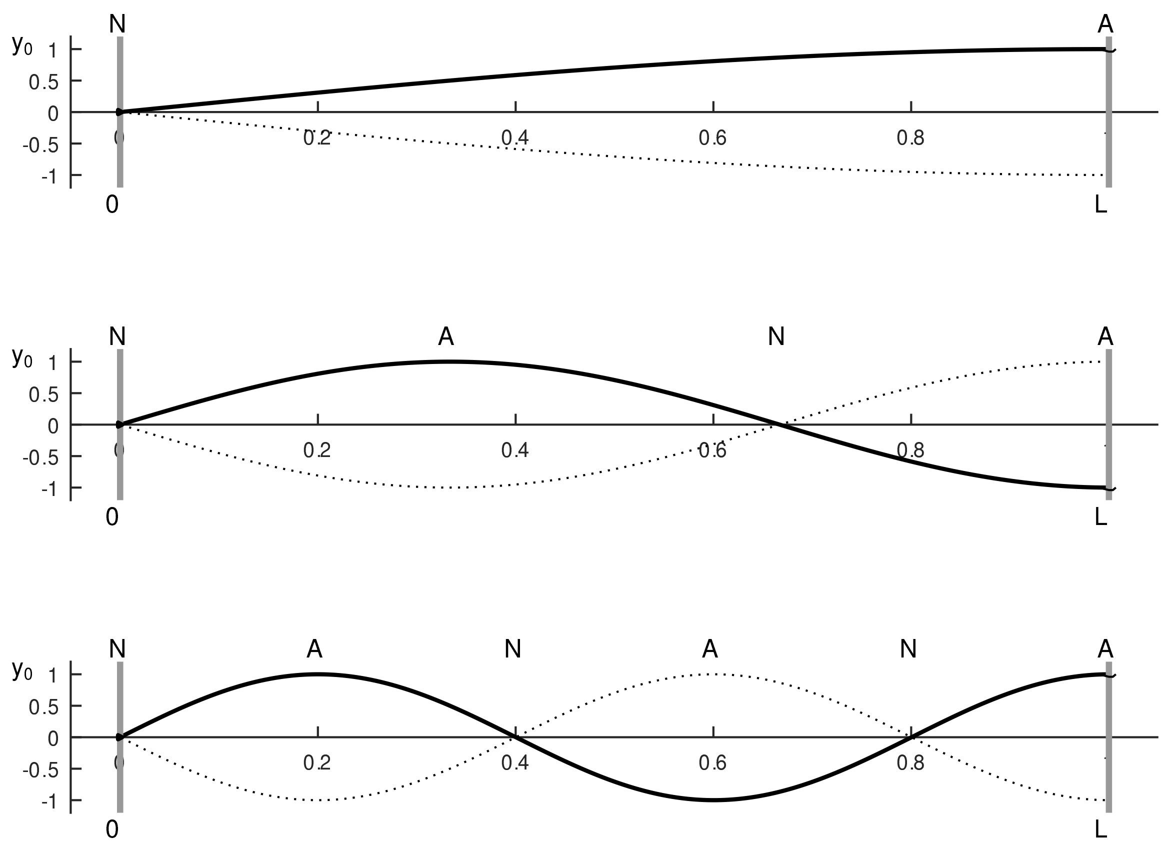


Figure 10.10: Wave modes on a string fixed at $x = 0$ and free (to slide up and down a frictionless rod under tension) at $x = L$. Free and fixed is the same sequence of wavelengths but the figure is mirror reversed.

To help out, I plotted the first three modes in figure 10.10, with nodes and antinodes labeled as always. In the second picture, L contains only 3/4 of a wavelength, so $\lambda = 4L/3$. In the third, one full wavelength seems to be $4L/5$. We might *guess, based on sausages alone*, that $\lambda = 4L, 4L/3, 4L/5, 4L/7...$, but the numbers on the bottom are *no longer the mode index* counted from the lowest frequency.

Let's do the math and see if our sausage-y guess is correct. For a node at $x = 0$ and an antinode at $x = L$, we should choose:

$$y(x,t) = y_0 \sin(kx)\cos(\omega t) \tag{10.64}$$

which guarantees that there is a node at $x = 0$. To get an antinode at the other end we note that:

$$\sin(\theta) = \pm 1 \quad \text{when} \quad \theta = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, ... \quad \text{or} \quad \theta = \theta_n = \frac{n\pi}{2} \quad \text{for} \quad n = 1, 3, 5...$$

Note Well: Only ***odd integer multiples of $\pi/2$*** appear in this sequence!

$$y(L,t) = \pm y_0 = y_0 \sin(k_n L)\cos(\omega_n t) \tag{10.65}$$

or:

$$k_n L = \frac{2\pi}{\lambda_n} L = \frac{n\pi}{2} \qquad \Rightarrow \qquad \lambda_n = \frac{4L}{n} \quad \text{with} \quad n = 1, 3, 5... \tag{10.66}$$

exactly as expected, where ***only odd integer n appear in the series!***

At this point finding the frequencies etc. (using the formulas from the box) should be *easy*, but because there is a point I want to emphasize I'll go ahead and fill in the frequencies of the ***first three modes***:

$$f_1 = \frac{v}{\lambda_1} = \frac{v}{4L} = \boxed{\frac{1}{4L}\sqrt{\frac{\mathcal{T}}{\mu}}} \tag{10.67}$$

$$f_3 = \frac{v}{\lambda_3} = \frac{3v}{4L} = \boxed{\frac{3}{4L}\sqrt{\frac{\mathcal{T}}{\mu}} = 3f_1} \tag{10.68}$$

$$f_5 = \frac{v}{\lambda_5} = \frac{5v}{2L} = \boxed{\frac{5}{4L}\sqrt{\frac{\mathcal{T}}{\mu}} = 5f_1} \tag{10.69}$$

...

Note that the labels on the frequencies *match* those on the wavelengths – they are $n = 1, 3, 5...$, not $m = 1, 2, 3...$ as a "mode index". Of course we *could* use the mode index, noting that $n = 2m - 1$ produces the correct odd integer sequence, but what we have found are the first three *harmonics* of the string. Reread the short section above on the difference between mode count from the bottom (longest wavelength, smallest frequency, principle harmonic) and the *harmonic frequencies* of the standing waves, which are *always integer multiples of the fundamental frequency!*

This is pretty much the only thing that makes differentiating mode and harmonic tricky. For strings fixed or free at *both* ends, it doesn't much matter as the two indices are interchangable, but for a string fixed at one end and free at the other it does, as it supports only the odd harmonic multiples of the fundamental.

Note that if we swap the boundary conditions so that the $x = 0$ end is free and the $x = L$ end is fixed, *nothing really changes*. Sure, we then need to use $y_0 \cos(kx)$ for the spatial waveform, but it turns out the nodes at $x = L$ occur only where $\cos(kL) = 0$, which leads to *exactly the same sequence of wavelengths and frequencies*.

Of course it does! We could have guessed this from symmetry alone! We can change one presentation of the problem into the other by simply ***relabelling*** $0 \Leftrightarrow L$! This is basically the same thing as walking around the string and viewing the *one* string from the other side, so it could hardly change the resonant frequencies and wavelengths, only their description. But we already know that changing coordinates changes that...

The one thing we haven't looked at or talked much about is energy. It turns out that (as one might expect) *energy is really important* in wave theory. Let's see why.

10.7: Energy

Clearly a wave can carry energy from one place to another. A cable we are coiling is hung up on a piece of wood. We flip a pulse onto the wire, it runs down to the piece of wood and knocks the wire free. My lungs and larnyx create sound waves, and those waves (as we'll see next week) resonantly wiggle hair cells in your cochleas and thereby trigger neurons running into your brains in your ears far away. The sun releases nuclear energy, and a few minutes later that energy, propagated to earth as a light wave, creates sugar energy stored inside a plant that is still later released while we play basketball.

It turns out that *almost all energy gets around in the form of waves!* Since moving energy around seems to be important, perhaps we should figure out how waves manage it, starting with waves on a string!

10.7.1: Energy in Traveling Waves

We'll start by considering the energy in a string carrying a *travelling harmonic wave of known angular frequency* ω. This will give use at least insight into energy in wave pulses as well, if only because arbitrary wave pulses can be fourier decomposed in harmonic traveling waves and the energy of the series summed up. It will also preview our analysis of standing waves, following.

Consider a (differentially) small piece of string located at position x along the string, of length dx and mass $dm = \mu dx$, oscillating up and down as a traveling wave moves down the string to the (why not?) right. This chunk of string must have (differential) kinetic energy:

$$dK = \frac{1}{2}\, dm\, v_y^2 \tag{10.70}$$

where v_y is the transverse velocity of the string at position x at time t. If the string is carrying a simple harmonic traveling wave to the right, then:

$$y(x,t) = y_0 \sin(kx - \omega t) \qquad v_y(x,t) = \frac{dy}{dt} = -\omega y_0 \cos(kx - \omega t)$$

and:

$$\boxed{dK = \frac{1}{2}\mu\omega^2 y_0^2 \cos^2(kx - \omega t)dx} \tag{10.71}$$

This is difficult to integrate across an *arbitrary* length of string, but it is actually ***very easy*** to integrate over any quarter-wavelength chunk of string! We'll choose to integrate it over a single wavelength, which (when we divide out the wavelength itself) will give us not the total kinetic energy on the string (which could be very long!) but rather the kinetic energy per wavelength, a quantity that will not vary (as it turns out) in time.

Here's how it goes. Using:

$$\theta = kx - \omega t \quad \Rightarrow \quad d\theta = k\,dx$$

we have to convert:

$$dx \Rightarrow \frac{1}{k}\,kdx \quad \text{and} \quad k\lambda \Rightarrow 2\pi$$

in the integral, using u (really θ in the present case) substitution:

$$\begin{aligned}
K_\lambda = \int_0^{K_\lambda} dK &= \int_0^\lambda \frac{1}{2}\mu\omega^2 y_0^2 \cos^2(kx - \omega t)\,dx \\
&= \frac{1}{2}\mu\omega^2 y_0^2 \frac{1}{k}\int_0^\lambda \cos^2(kx - \omega t)\,kdx \\
&= \frac{1}{2}\mu\omega^2 y_0^2 \frac{\lambda}{2\pi}\int_0^{2\pi} \cos^2(\theta)\,d\theta \\
&= \frac{1}{2}\mu\omega^2 y_0^2 \frac{\lambda}{2\pi} \times \pi \\
&= \frac{1}{4}\mu\omega^2 y_0^2 \lambda
\end{aligned} \tag{10.72}$$

where we have used the easily proven relation:

$$\int_0^{2\pi} \sin^2(\theta)d\theta = \int_0^{2\pi} \cos^2(\theta)d\theta = \pi \tag{10.73}$$

to do the final form of the integral.

The kinetic energy per wavelength is then:

$$\boxed{\frac{K_\lambda}{\lambda} = \frac{1}{4}\mu\omega^2 y_0^2} \tag{10.74}$$

If we assume that the string has length $L \gg \lambda$ the total kinetic energy it carries will be approximately:

$$K_{\text{tot}} \approx L \times \frac{K_\lambda}{\lambda} \tag{10.75}$$

In the limit of an infinitely long string, or a length of string that contains an integer number of wavelengths, this expression is exact; otherwise it might be off by the energy of a "remainder" fraction of a wavelength somewhere. The kinetic energy per wavelength is essentially the average energy density of the wave on the string. Note also that this answer *does not depend on time* as it was integrated out as an "irrelevant phase" at any instant in time along with x–

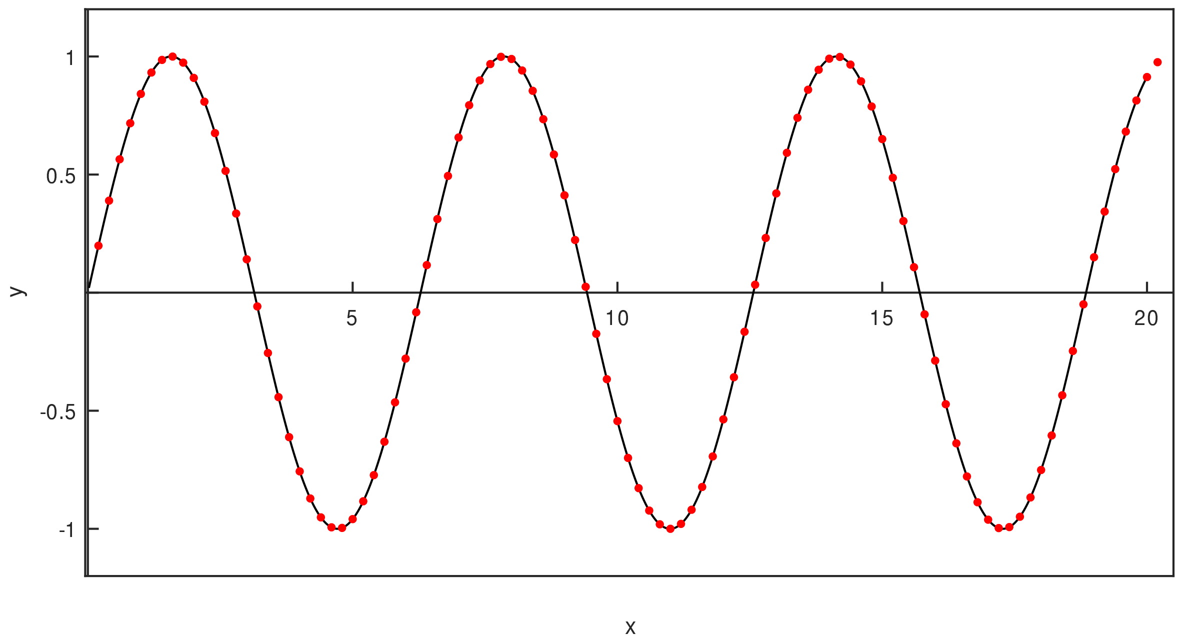


Figure 10.11: A few cycles of a traveling wave with red "beads" on the string that move up and down with the string strictly vertically so you can see the relative stretching of the string in between.

the total kinetic energy in a full wavelength does not change, although parts of the string are moving and parts are at momentary rest as the wave sweeps by.

The average potential energy of the string is more difficult. Where could the string be hiding potential energy? What is doing work on what as the wave is put onto the string and then propagates down it to be "received" by the end? Literally, if the string is stretched between two cans turned into a ***tin-can telephone***[262] .

In the last chapter we learned that putting pretty much *anything* under tension costs work as we "stretch" all of the little spring-like molecular bonds. Indeed, when we put some length of string under tension, it gets slightly longer and we do work getting it there. When there is no wave on it (and neglect gravity etc), it of course stretches out into a perfectly straight line.

Then we *bend* a wave form onto the string. The length of a sine curve (measured along the curve itself) is *strictly longer* than the x axis it sits on, because a straight line is the shortest distance between two points! One *has to do work* stretching the string against the tension $\mathcal{T}$ to put the sine wave onto the string, and that work is stored as potential energy *in the stretched string!* Technically, in all of the molecular bonds that hold the string together as a string.

If you explored the PHET simulation linked earlier in this chapter, you can see that this is how they modeled the string – a bunch of little masses *separated* by stretchable "springs", where we assume(d) at the beginning of the chapter that the chunks of string move only

[262]Wikipedia: http://www.wikipedia.org/wiki/Tin can telephone. Yes, this has its own web page. If you never made one of these and played with it as a young human, well, it is simple to make out and it's never too late...

vertically. This means that the string stretches the most where it has the greatest slope, and doesn't stretch at all where the slope is zero! I illustrate this in figure 10.11.

We need to use calculus to do better than one can do with small balls and springs. Just how much does a chunk of string with unstretched length dx stretch as it swings up and down? We can deduce the answer from figure 10.12. A small (eventually differentially small) chunk of the string of unstretched length Δx is seen to be stretched to a new length $\Delta \ell = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ when it is moved straight up to the height $y(x,t)$ so that it has some slope there.

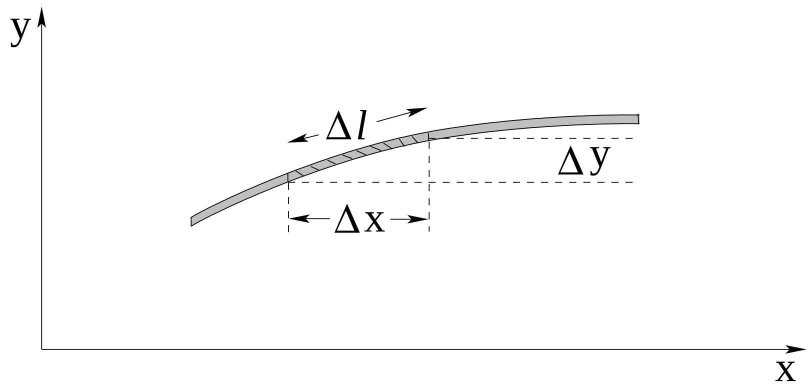


Figure 10.12: A small chunk of the string. The chunk that (unstretched at $y(x,t) = 0$ has length Δx is stretched by an amount $\Delta \ell - \Delta x$.

If we assume that this slope is very small ($\Delta y \ll \Delta x$, consistent with our small angle approximation) then:

$$\begin{aligned}
\Delta \ell - \Delta x &= \left((\Delta x)^2 + (\Delta y)^2\right)^{1/2} - \Delta x \\
&= \Delta x \left(1 + \frac{(\Delta y)^2}{(\Delta x)^2}\right)^{1/2} - \Delta x \\
&= \Delta x \left(1 + \frac{1}{2}\left(\frac{\Delta y}{\Delta x}\right)^2 + ...\right) - \Delta x \\
&\approx \frac{\Delta x}{2}\left(\frac{\Delta y}{\Delta x}\right)^2
\end{aligned} \tag{10.76}$$

In the limit that $\Delta x \to 0$, this becomes:

$$d\ell - dx = \frac{1}{2}\left(\frac{dy}{dx}\right)^2 dx \tag{10.77}$$

The work done stretching this small chunk of string to its new length against the (nearly constant) tension $\mathcal{T}$ is the differential potential energy of the chunk of string of differential length dx:

$$dU = T(d\ell - dx) = \frac{T}{2}\left(\frac{dy}{dx}\right)^2 dx \tag{10.78}$$

As before, we need to integrate this to find the potential energy in a single wavelength. For a *travelling harmonic wave* $y(x,t) = y_0 \sin(kx - \omega t)$ we considered before:

$$\frac{dy}{dx} = k y_0 \cos(kx - \omega t) \tag{10.79}$$

and hence:

$$dU = \frac{1}{2} T k^2 y_0^2 \cos^2(kx - \omega t) dx \tag{10.80}$$

Before integrating this, let's multiply by $\frac{\mu}{\mu} = 1$ and do a bit of algebraic rearrangement:

$$\begin{aligned} dU &= \frac{\mu}{\mu} \times \frac{1}{2} T k^2 y_0^2 \cos^2(kx - \omega t) dx \\ &= \frac{1}{2} \mu k^2 \frac{T}{\mu} y_0^2 \cos^2(kx - \omega t) dx \\ &= \frac{1}{2} \mu k^2 v^2 y_0^2 \cos^2(kx - \omega t) dx \end{aligned}$$

or

$$\boxed{dU = \frac{1}{2} \mu \omega^2 y_0^2 \cos^2(kx - \omega t) dx} \tag{10.81}$$

But this is *absolutely identical to dK!* We have saved ourselves from having to do the integral a second time – the result is just:

$$\boxed{\frac{U_\lambda}{\lambda} = \frac{1}{4} \mu \omega^2 y_0^2} \tag{10.82}$$

If we combine the kinetic and potential energy densities (dK/dx and dU/dx) we get the *total linear energy density* in the traveling wave on the string:

$$\frac{dE}{dx} = \frac{dK}{dx} + \frac{dU}{dx} = \mu \omega^2 y_0^2 \cos^2(kx - \omega t) \tag{10.83}$$

Note well that the energy density itself is a traveling wave – energy is *carried down the string* as the part of the string that is both maximally stretched *and* moving fastest (both occurring where the wave has the greatest slope) moves to the right at speed v.

It is worth pointing out that this is *different* from the behavior we observed for a *simple* harmonic oscillator, where the potential and kinetic energy are also equal but *out of phase*. The string oscillates, but it does so because the stretching in the string propagates as a longitudinal wave *along* the string (mostly in the x direction) but it acts at an angle so there is always a net force lifting up or pulling down the the string in the vertical (y) direction!

Finally, the total energy per wavelength is also useful as a good approximation to the average energy per unit length on the string as long as the string is long enough to contain many wavelengths.

$$\boxed{\frac{E_\lambda}{\lambda} = \frac{1}{2} \mu \omega^2 y_0^2} \tag{10.84}$$

This expression is important. Energy carried by sound waves and light waves have *very similar expressions* for their energy density, in particular always being proportional to harmonic wave amplitude squared.

10.7.2: Power Transmission by a Travelling Wave

We haven't discussed how to put a travelling wave *onto* a string or how to take it *off* at the other end, but in a nutshell, we put it on by doing *work* on (say) the left hand end of the string

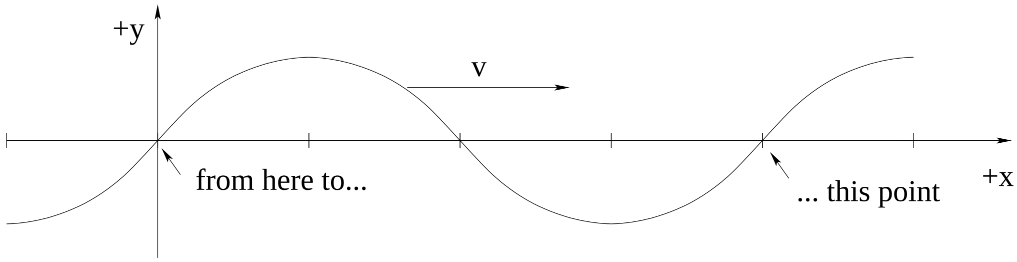


Figure 10.13: *All* of the energy in a single wavelength of the wave is transmitted past an (arbitrary) point on the string in the direction of motion in *one* period of oscillation.

and letting the string do work on something attached to the right hand end of the string. In that way, energy added at the left can be removed and used at the right. Waves transmit energy[263]!

Figure 10.13 above gives us a simple model of energy transmission down a string that is nearly universal to wave theory (and not just waves on strings – sound waves or electromagnetic waves have a *similar* picture, but in more dimensions). A harmonic travelling wave going from left to right on a string:

$$y(x,t) = y_0 \sin(kx - \omega t)$$

has (as shown in the previous section) a total energy in one wavelength given by:

$$\Delta E = E_\lambda = \frac{1}{2}\mu\omega^2 y_0^2 \lambda \tag{10.85}$$

All of this energy moves from left to right along with the wave. In one period, *all* of the energy in one wavelength passes any given point on the string, so the *average* energy per unit time passing that point is:

$$P = \frac{\Delta E}{\Delta t} = \frac{1}{2}\mu\omega^2 y_0^2 \frac{\lambda}{T} = \frac{1}{2}\mu\omega^2 y_0^2 v \tag{10.86}$$

In words, the average power transmitted by the string is the energy density times the speed of the travelling wave on the string. Since we will almost always be more interested in average power transmitted down the string than in the instantaneous rate work is being done at the end points of the string where it connects to a receiver of some sort, we will always use this expression and usually omit the qualifier "average" as being understood.

10.7.3: Energy in Standing Waves

Next, Consider a generic *standing wave* $y(x,t) = y_0 \sin(kx)\cos(\omega t)$ for a string that (perhaps) is fixed at ends located at $x = 0$ and $x = L$. We can easily evaluate the kinetic energy of a chunk of length dx:

$$dK = \frac{1}{2}\mu\omega^2 y_0^2 \sin^2(kx)\sin^2(\omega t)dx \tag{10.87}$$

and the potential energy of the same chunk:

$$\begin{aligned} dU &= \frac{1}{2}Tk^2 y_0^2 \cos^2(kx)\cos^2(\omega t)dx \\ &= \frac{1}{2}\mu\omega^2 y_0^2 \cos^2(kx)\cos^2(\omega t)dx \end{aligned} \tag{10.88}$$

[263]A reminder. All it takes is two plastic cups with a long piece of fishing line attached to the ends stretched in between them. Speak into one of them while your friend holds the other to their ear! You too can transmit waves down a string that encode your voice, and you don't even need tin cans!

If we now form the total energy density of this chunk of string as before:

$$\frac{dE}{dx} = \frac{1}{2}\mu\omega^2 y_0^2 \left(\sin^2(kx)\sin^2(\omega t) + \cos^2(kx)\cos^2(\omega t)\right) \tag{10.89}$$

we note that the energy density of the string is *not a constant*.

At first this appears to be a problem, but really it is not. All this tells us is that the nodes and antinodes of the string (where $\sin(kx)$ or $\cos(kx)$ are zero, respectively) carry the energy of the string *out of phase* with one another. At times $\omega t = 0, \pi, 2\pi...$, the potential energy is maximal, the kinetic energy is zero, and the potential energy is concentrated at the points where $\cos(kx) = \pm 1$. These are the *nodes*, which have maximum magnitude of slope and stretch as the string reaches its maximum positive or negative amplitude. At times $\omega t = \pi/2, 3\pi/2...$ the kinetic energy is maximal, the potential energy is zero (indeed, the string is now flat and unstretched at all) and the kinetic energy is concentrated at the *antinodes*, where the velocity of the string is the greatest.

This makes perfect sense! The energy in the string oscillates from being all potential (for the entire string) as it momentarily comes to rest to all kinetic as the string is momentarily straight and moving at maximum speed (for the location) everywhere. The one question we need to answer, though, is: Is the *total* energy of the string constant? We would be sad if it were not, because we know that the endpoints of the standing wave are (for example) fixed, and hence no work is done there. Whatever the energy in the wave on the string, it has nowhere to go.

We note that no matter what the boundary conditions, the string contains an integer number of quarter wavelengths. All we have to do is show that the total energy in any quarter wavelength of the string is constant in time and we're good to go. Consider the following integral:

$$\begin{aligned}\int_0^{\lambda/4} \cos^2(kx)dx = \int_0^{\lambda/4} \sin^2(kx)dx &= \frac{1}{k}\int_0^{\lambda/4} \sin^2(kx)k dx \\ &= \frac{1}{k}\int_0^{\pi/2} sin^2(\theta)d\theta \\ &= \frac{\lambda}{2\pi}\frac{\pi}{4} = \frac{\lambda}{8}\end{aligned} \tag{10.90}$$

(where $\sin^2$ and $\cos^2$ have the same integral from symmetry, one equals the other with the addition or subtraction of the constant angle $\pi/2$ which doesn't change the integral). Note also that it doesn't matter if we are doing quarter wavelengths or quarter periods of a time-dependent harmonic wave:

$$\int_0^{T/4} \cos^2(\omega t)dt = \int_0^{T/4} \sin^2(\omega t)dt = \frac{T}{8} \tag{10.91}$$

The time average of these quantities (over any integer number of quarter periods or quarter wavelengths) are thus always $\frac{1}{2}$:

$$\frac{4}{T}\int_0^{T/4} \cos^2(\omega t)dt = \frac{4}{T}\int_0^{T/4} \sin^2(\omega t)dt = \frac{4}{T}]\frac{T}{8} = \frac{1}{2} \tag{10.92}$$

Finally, if we let the time or length go to ∞ to do our time or space average, we always asymptotically approach $\frac{1}{2}$ because even if there is a non-quarter wavelength fraction of a wave leftover at one end (the remainder after summing all the full quarter waves) its contribution vanishes as time (or distance) goes to infinity.

We summarize as follows:

The time or space average of any function that goes
like $\sin^2$ ***or*** $\cos^2$ ***(of*** ωt ***or*** kx ***respectively) is one half!***

This is worth remembering, as we use it quite often.

From this we can easily find the energy per wavelength of any standing wave on a string of length L (which as noted has an integer number of quarter waves on it). Expressing it this way also facilitates comparison with a travelling wave (where we obtained the energy per wavelength)

$$\begin{aligned}
\frac{E_\lambda}{\lambda} &= \frac{1}{\lambda}\frac{1}{2}\mu\omega^2 y_0^2 \left(\sin^2(\omega t)\int_0^\lambda \sin^2(kx)dx + \cos^2(\omega t)\int_0^\lambda \cos^2(kx)\right) \\
&= \frac{1}{\lambda}\frac{1}{2}\mu\omega^2 y_0^2 \left(\sin^2(\omega t)\frac{\lambda}{2} + \cos^2(\omega t)\frac{\lambda}{2}\right) \\
&= \frac{1}{4}\mu\omega^2 y_0^2 \left(\sin^2(\omega t) + \cos^2(\omega t)\right) \\
&= \frac{1}{4}\mu\omega^2 y_0^2 \qquad (10.93)
\end{aligned}$$

From this we can easily find the total energy on the string. Up to here we haven't bothered to specify what the string boundary conditions are because we worked with quarter wavelengths in the integral, which works for any of them. Suppose n is the harmonic index of a string with two fixed *or* free ends, so $\lambda_n = 2L/n$, and y_n is its amplitude. The energy in a single wavelength can then be written in terms of L as:

$$E_{\lambda_n} = \frac{n}{4}\mu\omega_n^2 y_n^2 \lambda_n = \frac{1}{4}\mu\omega_n^2 y_n^2 \left(\frac{2L}{n}\right) \qquad (10.94)$$

But there are also $n/2$ wavelengths in the total length L of the string! So the *total energy in the string of length* L *is*:

$$E_{\rm tot} = \frac{1}{4}\mu\omega_n^2 y_n^2 \left(\frac{2L}{n}\right) \times \frac{n}{2} = \frac{1}{4}\mu\omega_n^2 y_n^2 L \qquad (10.95)$$

This result is exact, and depends on n per se only via ω_n. It is left as an exercise to show that it is equally true for mixed boundary conditions and the consequent odd harmonics. The energy in a wavelength divided by the wavelength is the average energy per unit length for *both* traveling waves *and* stationary waves.

The last thing to talk about before going on is understanding why the expression for the energy in a wavelength has a factor of $1/4$ for a standing wave where the otherwise identical expression for energy in wavelength in a traveling wave hadn $1/2$. After all, didn't we add *two* waves to make a stationary wave out of traveling waves?

We can understand this easily enough. The standing wave consists of two travelling waves, each with *half of the amplitude* y_0 *of the standing wave*, going in opposite directions. As we show in the next section, the total energies per wavelength in these two waves are *independent* – we just add them! Energy per wavelength is proportional to amplitude squared, so each traveling wave with amplitude $y_0/2$ has $1/4$ of the energy of a travelling wave with amplitude y_0. But there are two of them, hence an *extra* factor of $1/2$, making the factor for standing waves $1/2 \times 1/2 = 1/4$. Simple enough!

Another way of thinking about it is that for a traveling wave, the entire string oscillates harmonically through the amplitude y_0, and every bit of string has peak values for both U and K. For a standing wave, only antinodes oscillate through the whole K limit, only nodes oscillate through the whole U limit, on average they have *half of the energy per wavelength* as a consequence. Understanding this (eitehr or both ways) can help you remember which is which.

10.7.4: Energy of Wave Superpositions

We proved at the beginning of this section that we can sum any number of solutions to the 1DWE (scaled by their respective amplitudes) and the result will still satisfy the 1DWE. We *also* have seen how energy is expressed in the string as a sum of potential energy (where the string is stretched by the wave) and kinetic energy (where the string is moving). The only way – so far – for energy to get on or off the string is at the ends, where perhaps a non-conservative "hand" is shaking the string up and down to put the wave on it, or the other end is connected to a lever that is shaken up and down as the wave impinges on it.

Suppose we have just *2* distinct harmonic traveling waves:

$$y_1(x,t) = y_1 \sin(k_1 x - \omega_1 t) \tag{10.96}$$
$$y_2(x,t) = y_2 \sin(k_2 x - \omega_2 t) \tag{10.97}$$

where $k_1 \neq k_2$ – so they are "distinct" solutions to the 1DWE on an "infinite" string[264]. We build a solution that is the simple superposition of these 2 traveling waves (hopefully it is obvious that we could add 2, 3, 4 or more of them and get identical results):

$$y(x,t) = y_1 \sin(k_1 x - \omega_1 t) + y_2 \sin(k_2 x - \omega_2 t) \tag{10.98}$$

We can ask: what is the (average) energy density of waves on this string?

To answer this, we have to evaluate (three) integral(s):

$$\begin{aligned}\left(\frac{dE}{dx}\right)_{\text{avg}} = \frac{1}{2L}\int_{-L}^{L} \mu\omega_1^2 y_1^2 \cos^2(k_x x - \omega_1 t)\,dx \\ + \frac{1}{2L}\int_{-L}^{L} \mu \times 2\omega_1\omega_2 y_1 y_2 \cos(k_1 x - \omega_1 t)\cos(k_2 x - \omega_2 t)\,dx \\ + \frac{1}{2L}\int_{-L}^{L} \mu\omega_2^2 y_2^2 \cos^2(k_2 x - \omega_2 t)\,dx\end{aligned} \tag{10.99}$$

in the limit that $L \to \infty$.

The two integrals over $\cos^2(k_{1,2}x - \omega_{1,2}t)$ are "easy", because we've really already done

[264]Or at least, one that is much longer than either of their wavelengths.

them! Integrating over any *single wavelength* gives:

$$\frac{E_1}{\lambda_1} = \frac{1}{2}\mu\omega_1^2 y_1^2 \tag{10.100}$$

$$\frac{E_2}{\lambda_2} = \frac{1}{2}\mu\omega_2^2 y_2^2 \tag{10.101}$$

If $L \gg \lambda_{1,2}$, the *average* energy per unit length just equals the energy per wavelength, as discussed earlier.

What remains is integrating:

$$\frac{1}{2L}\int_{-L}^{L} 2\mu\omega_1\omega_2 y_1 y_2 \cos(k_1x - \omega_1 t)\cos(k_2x - \omega_2 t)dx$$
$$= 2\mu\omega_1\omega_2 y_1 y_2 \frac{1}{2L}\int_{-L}^{L} \cos(k_1x - \omega_1 t)\cos(k_2x - \omega_2 t)dx \tag{10.102}$$

where $\boldsymbol{k_1 \neq k_2}$. To evaluate this, we'll use the trig identity:

$$\cos(A)\cos(B) = \frac{1}{2}\left\{\cos(A+B) + \cos(A-B)\}\right)$$

The integral (only) becomes:

$$\frac{1}{2L}\int_{-L}^{L} \cos((k_1+k_2)x - (\omega_1 + \omega_2 t) + \frac{1}{2L}\int_{-L}^{L} \cos((k1-k_2)x - (\omega_1 - \omega_2)t)dx \tag{10.103}$$

As long as $k_1 \neq k_2$, these integrals in the limit of very large L are both ***zero!*** The integral of *any* sinusoidal function over *any* interval is never larger than 1, so that the integrals must vanish like at least:

$$\lim_{L\to\infty} \frac{1}{2L} = 0$$

In other words, no matter how many traveling wave solutions we include, and no matter what direction they travel, as long as they have distinct k_i wave numbers the cross terms in the average energy density in some length $L \to \infty$ *vanish*, while the *same*-k terms each contribute the average energy density of that particular wave only.

This is sufficient to show that the average energy density of the string is the sum of the average energy densities of all of the waves, which in turn are just their *independent* energies per wavelength!

$$\left(\frac{dE}{dx}\right)_{\rm avg} = \frac{E_1}{\lambda_1} + \frac{E_2}{\lambda_2}... + \frac{E_N}{\lambda_N} = \sum_i^N \frac{E_i}{\lambda_i}$$

or:

$$\boxed{\left(\frac{dE}{dx}\right)_{\rm avg} = \sum_i^N \frac{1}{2}\mu\omega_i^2 y_i^2} \tag{10.104}$$

This relation expresses energy conservation for a wave built of a discrete superposition of waves with different k, and is an entree into building a wave out of a *continuous integral* of waves *over* k with the amplitude a continuous function of k, an exercise most math and physics majors will very likely undertake one day[265] . The total energy of the resulting waveform is the

[265]Wikipedia: http://www.wikipedia.org/wiki/Fourier Transform. But today is not the day, unless you choose to bite off the weighty chunk in this wikipedia page. This page is a bit math-y-er than I care for, actually – one has to use and understand complex exponential waveforms from the beginning. Sorry.

simple sum of the energy in each waveform! Note that for traveling waves evaluated on *short* chunks of string this will not work, because our energy per unit length only had a nice form if that length was (an integer multiple of) a wavelength and because the integrals over the cross terms will *not* necessarily be small until L is much greater than the largest wavelength in the superposition.

This dual result – that energy is "conserved" as waves move down a string, and that the energy density is the sum of the energy densities of each unique contributing harmonic wave in a superposition – is pretty important, and we implicitly used parts of the reasoning involved when we analyzed wave transmission and reflection for wave pulses and motivated standing waves.

For standing waves it is actually a lot simpler. There, a given string only supports a *discrete set of wave modes* that satisfy the boundary conditions. If we imagine a superposition of only two allowed modes on a string fixed at both ends, for example (again, expecting to be able to generalize our result "by inspection" to include more modes).

Suppose the two modes have *harmonic* index n and m. Then:

$$y(x,t) = y_n \sin(k_n x)\cos(\omega_n t) + y_m \sin(k_m x)\cos(\omega_m t) \tag{10.105}$$

where $n \neq m$ and:

$$k_n = \frac{n\pi}{L} \qquad k_m = \frac{m\pi}{L}$$

then:

$$\frac{dE}{dx} = \frac{1}{2}\mu\left(\frac{dy(x,t)}{dt}\right)^2 + \frac{1}{2}T\left(\frac{dy(x,t)}{dx}\right)^2 \tag{10.106}$$

where:

$$\frac{dy(x,t)}{dt} = -\omega_n y_n \sin(k_n x)\sin(\omega_n t) - \omega_m y_m \sin(k_m x)\sin(\omega_m t) \tag{10.107}$$

$$\frac{dy(x,t)}{dx} = k_n y_n \cos(k_n x)\cos(\omega_n t) + k_m y_m \cos(k_m x)\cos(\omega_m t) \tag{10.108}$$

Squaring these leads to a bit of a mess. However, we don't really care about the constants – all that really matters is the square of the trig function parts. Those come in four combinations for each derivative, and hopefully it is obvious that we'll get the same *result* from integrating either the $\sin$ or $\cos$ squared forms. Similarly, only two are really distinct, since n and m are arbitrary integer harmonic indices greater than or equal to 1, we can freely relabel then and nothing will change. We therefore only have to look carefully at *two possible forms* to figure out the whole thing. They are:

$$\sin^2(k_n x)\sin^2(\omega_n t) \quad \text{and} \quad \sin(k_n x)\sin(k_m x)\sin(\omega_n t)\sin(\omega_m t)$$

The first one is easy. We change variables for

$$\int_0^L \sin^2(k_n x)\sin^2(\omega_n t)\,dx \tag{10.109}$$

to $\theta = k_n x$ (multiplying and dividing by k_n and converting the function, the limits and the differential to θ form):

$$\frac{\lambda_n}{2\pi}\int_0^{n\pi} \sin^2(k_n x)\sin^2(\omega_n t)\,k_n dx = n\frac{\lambda_n}{4}\sin^2(\omega_n t) \tag{10.110}$$

(and exactly the same result, shifted in phase):

$$\frac{\lambda_n}{2\pi}\int_0^{n\pi}\cos^2(k_n x)\cos^2(\omega_n t)\,k_n dx = n\frac{\lambda_n}{4}\cos^2(\omega_n t) \tag{10.111}$$

Adding them yields the same result we got above for the energy per wavelength of a single standing wave mode, summed over the (integer) number of wavelengths on the string. The average energy density for the square of the n and m terms directly is exactly what we got above, summed:

$$\frac{1}{4}\mu\left(\omega_n^2 y_n^2 + \omega_m^2 y_m^2\right) \tag{10.112}$$

But what about the cross term?

$$\sin(\omega_n t)\sin(\omega_m t)\int_0^L \sin(k_n x)\sin(k_m x)\,dx \tag{10.113}$$

Here it is probably simplest to substitute $k_n x = n\pi x/L$, $k_m x = m\pi x/L$ before using another trig identity:

$$\sin(A)\sin(B) = \frac{1}{2}\left\{\cos(A - B - \cos(A + B)\right\}$$

This makes the integral we must do into:

$$\begin{aligned}\int_0^L \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)\,dx = \\ \frac{1}{2}\left\{\int_0^L \cos\left(\frac{(n-m)\pi x}{L}\right) - \int_0^L \cos\left(\frac{(n+m)\pi x}{L}\right)\right\}\,dx \\ \frac{1}{2}\left\{\int_0^{(n-m)\pi}\cos(\theta)\,d\theta - \int_0^{(n+m)\pi}\cos(\theta)\,d\theta\right\}\end{aligned} \tag{10.114}$$

If $n - m$ is odd, the $n + m$ is odd as well, in which case the integrals are both either 1 or -1 and the total is zero. If $n - m$ is even, $n + m$ is even as well and both integrals are zero. No matter how you slice it, the answer is zero[266]:

$$\int_0^L \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)\,dx = 0 \quad \text{if integer } n \neq m \tag{10.115}$$

In other words, the cross-terms are zero, and – as we might expect, based on our study of traveling vs standing waves so far. Note that exactly the same thing holds for cosine instead of sine.

This is sufficient for us to find the total energy of an arbitrary superposition of standing wave modes on a string fixed or free at either end in any permutation. The energies of the modes simply add with no interference between modes. Using equation 10.95:

$$\boxed{E_{\text{tot}} = \sum_n \frac{1}{4}\mu\omega_n^2 y_n^2 L} \tag{10.116}$$

This actually makes sense. The *work* required to put energy into any given mode is independent of that for any other mode, and energy put onto a string has nowhere to go in this "ideal"

[266]There are a number of ways of arriving at this answer – I just picked one that is similar to our argument for traveling waves.

model where we ignore damping. Hence the total energy is just the sum of the individual energies in each mode in the superposition.

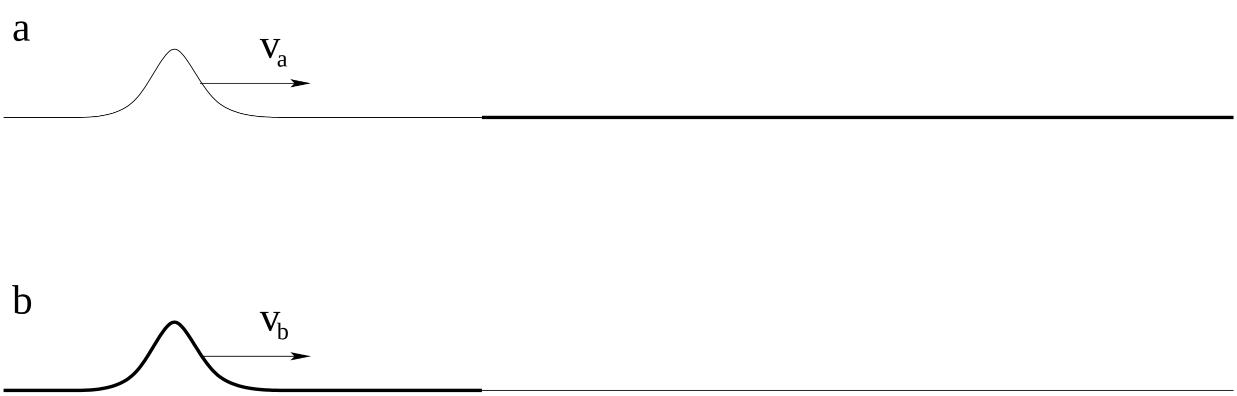
This is a *very important result* in wave theory, and more elegant forms of the integrals we did along the way form the basis of functional analysis – representing arbitrary 1D functions as superpositions of e.g. discrete harmonic waves in the case of Fourier Series or integrals over continues harmonic wave distributions in the case of the Fourier Transform. Both are beyond the scope of this course to discuss further, but form an important part of the future pathway forward for serious physics, math or engineering students. Hopefully this abbreviated discussion will get you off on the right foot if/when you take classes that cover this.

Homework for Week 10

Problem 1.

Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

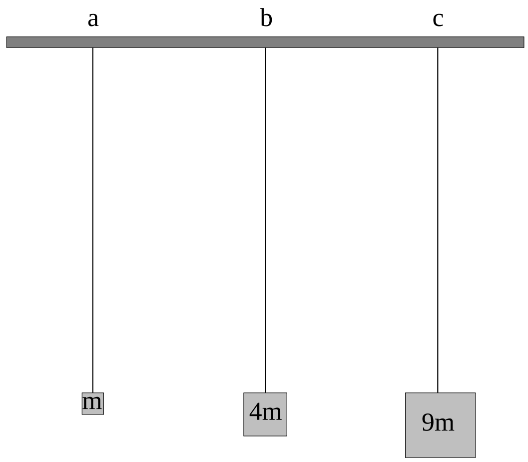
Problem 2.



Two combinations of two strings with different mass densities are drawn above that are connected at a smooth junction as shown. In both cases the string with the greatest mass density is drawn ***darker*** and ***thicker*** than the lighter one, and the strings have the *same tension* T in both a and b. A wave pulse is generated on the string pairs that is travelling from left to right as shown. The wave pulse will arrive at the junction between the strings at time t_a (for a) and t_b (for b).

Sketch reasonable *estimates* for the ***transmitted and reflected wave pulses*** onto (copies of) the a and b figures at time $2t_a$ and $2t_b$ respectively. Your sketch should correctly represent things like the ***relative speed of the reflected and transmitted wave*** and any changes you might reasonably expect for the ***amplitude and appearance*** of the pulses.

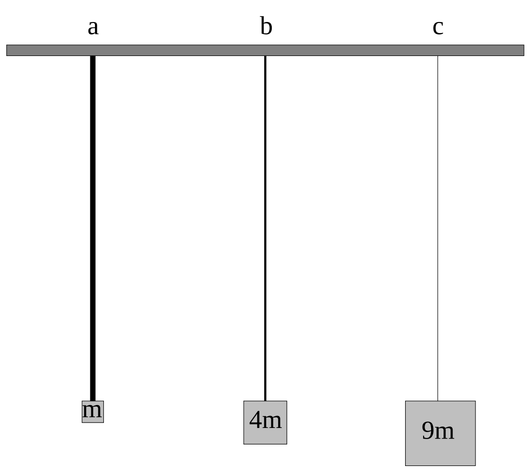
Problem 3.



i	v_i
a	$v_a = \boxed{1} \times v_0$
b	$v_b = \boxed{} \times v_0$
c	$v_c = \boxed{} \times v_0$

Three strings of length L (not shown) with the same mass per unit length μ are suspended vertically and blocks of mass m, $4m$ and $9m$ are hung from them. The total mass of each string $\mu L \ll m$ (the strings are *much* lighter than the masses hanging from them). If the speed of a wave pulse on the first string is $v_a = v_0$, fill in the provided table (on your own paper, of course) with entries for v_b and v_c:

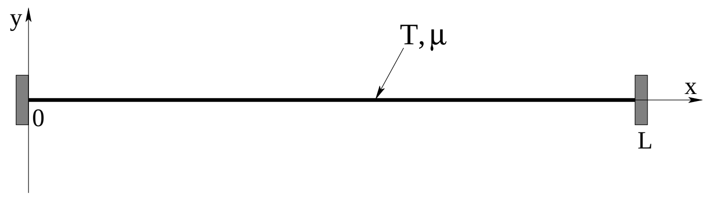
Problem 4.



i	v_i
a	$v_a = \boxed{1} \times v_0$
b	$v_b = \boxed{} \times v_0$
c	$v_c = \boxed{} \times v_0$

Three strings of length L have *different* mass per unit length and *different* masses hanging from them. String a has mass per unit length μ_a and a mass mg hanging from it, and has speed $v_a = v_0$. Similarly, $\mu_b = \frac{1}{2}\mu_a$ and has $4mg$ hanging from it and $\mu_c = \frac{1}{4}\mu_a$ with $9mg$ hanging. In all three cases, the total weight of the string is ***much less*** than the weight hanging from it. Fill in the provided table (on your own paper, of course) with entries for v_b and v_c:

Problem 5.



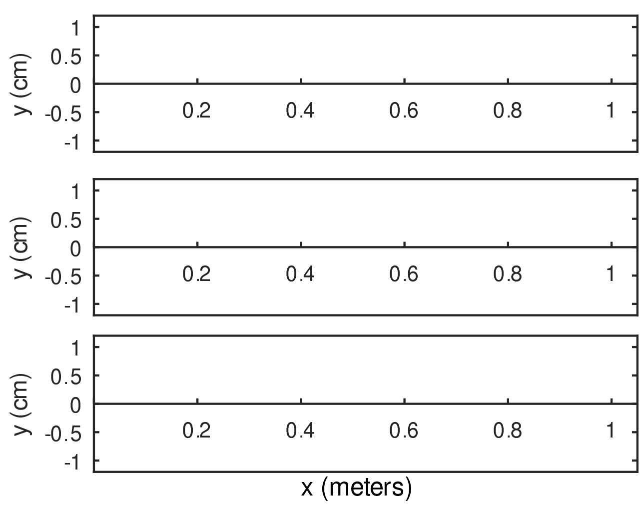
A string of mass density $\mu = 0.001$ kg/m is stretched to a tension $T = 10$ N and is ***fixed*** at ***both*** $\boldsymbol{x = 0}$ and $\boldsymbol{x = L = 1}$ meter. All answers should be given in terms of these quantities or new quantities you define in terms of these quantities, and may contain symbols like $\pi, \sqrt{}$.

a) What is the speed of waves on this string?

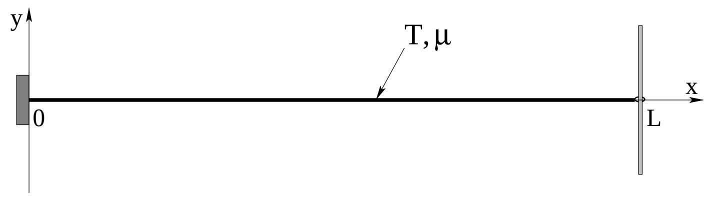
b) Write down the equation $y_n(x, t)$ for a generic ***standing wave*** on this string with ***harmonic*** index n in terms of ω_n and k_n, assuming that the string is maximally displaced at $t = 0$. Verify that it is a solution to the 1D wave equation for waves on the string. Remember that the string is fixed at both ends!

c) Find $k_n, \omega_n, f_n, \lambda_n$ for the ***first three modes supported by the string*** – the answers will all be numbers, with units, and may contain symbols such as π.

d) Sketch these three modes in on (a copy of) the axes below for an amplitude $A = 1$ cm. Note that you should be able to find at least the wavelengths from the pictures alone! Place a star next to the graph of the ***third harmonic***.



Problem 6.



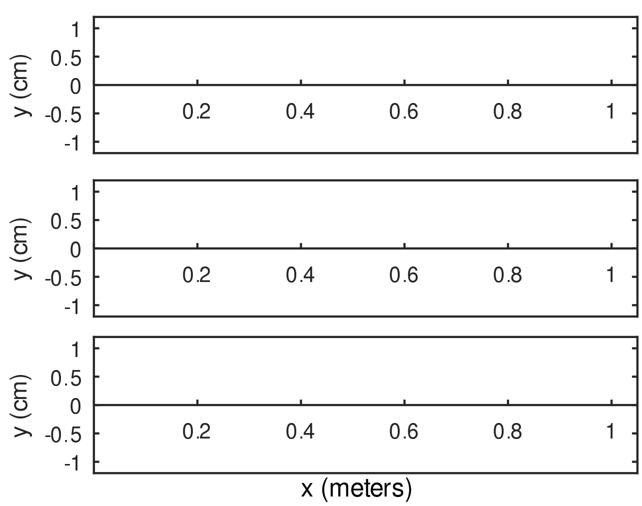
A string of mass density $\mu = 0.0005$ kg/m is stretched to a tension $T = 20$ N and is ***fixed at $x = 0$ and free (frictionless loop) at $x = L = 1$*** meter. All answers should be given in terms of these quantities or new quantities you define in terms of these quantities, and may contain symbols like $\pi, \sqrt{}$.

a) What is the speed of waves on this string?

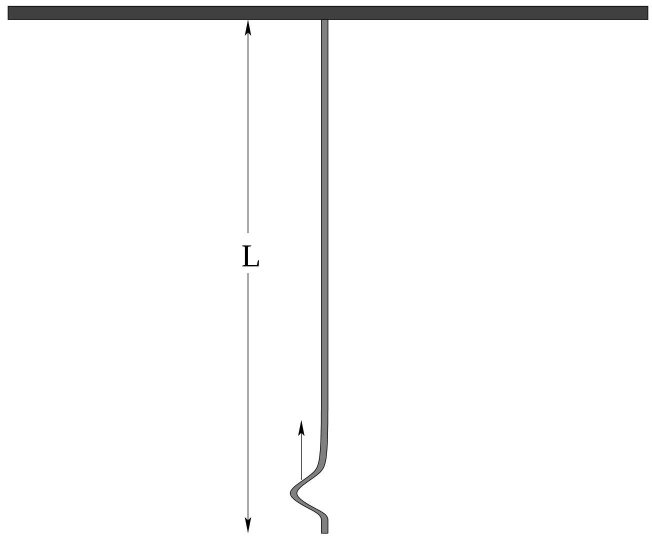
b) Write down the equation $y_n(x,t)$ for a generic ***standing wave*** on this string with ***harmonic*** index n in terms of ω_n and k_n, assuming that the string is maximally displaced at $t = 0$. Verify that it is a solution to the 1D wave equation for waves on the string. Remember that the string is fixed at one end and ***free*** at the other!

c) Find $k_n, \omega_n, f_n, \lambda_n$ for the ***first three modes supported by the string*** – the answers will all be numbers, with units, and may contain symbols such as π.

d) Sketch these three modes in on (a copy of) the axes below for an amplitude $A = 1$ cm. Note that you should be able to find at least the wavelengths from the pictures alone! Place a star next to the graph of the ***third harmonic***.



Problem 7.



A massive string with mass density μ and total length L is hanging from the ceiling as shown.

a) Find the tension $T(y)$ in the string as a function of y, the distance measured *up* from its bottom end. Hint: The string is *not massless*, so each small bit of string must be in ***static equilibrium***.

b) Find the velocity $v(y)$ of a small wave pulse cast into the string at the bottom that is travelling upward.

c) Use calculus to find the amount of time it will take this pulse to reach the top of the string, reflect, and return to the bottom. Neglect the size (width in y) of the pulse relative to the length of the string.

Problem 8.

You are given the following information resulting from measurements of the standing wave modes of two distinct strings, a) and b), both of length $L = 0.5$ meter, that have unknown boundary conditions. Use the information to answer the following questions

For the first string (a), you are told that two ***successive*** resonant frequencies are $f_i = 150$ Hz and $f_{i+1} = 210$ Hz where the ***mode*** index i counts the frequencies from the bottom, starting at $i = 1$ for the fundamental frequency. You are also told that the mass density for *this* string is $\mu_a = 1$ gram/meter. Find in the units indicated:

$$\text{(mode index) } i = \boxed{} \qquad \text{(fundamental frequency) } f_1 = \boxed{} \text{ Hz}$$

$$\text{(wave speed) } v_a = \boxed{} \text{ m/sec} \qquad \text{(tension) } T_a = \boxed{} \text{ N}$$

Select the possible boundary condition(s) for this string:

☐ Fixed at both ends. ☐ Free at both ends.

☐ Fixed at one end, free at the other.

For the second string (b), you are told that two ***successive*** resonant frequencies are $f_i = 60$ Hz and $f_{i+1} = 75$ Hz where the ***mode*** index i counts the frequencies from the bottom, starting at $i = 1$ for the fundamental frequency. You are also told that the tension for *this* string is $T_b = 10$ N. Find in the units indicated:

$$\text{(mode index) } i = \boxed{} \qquad \text{(fundamental frequency) } f_1 = \boxed{} \text{ Hz}$$

$$\text{(wave speed) } v_b = \boxed{} \text{ m/sec} \qquad \text{(mass density) } \mu_b = \boxed{} \text{ grams/meter}$$

Select the possible boundary condition(s) for this string:

☐ Fixed at both ends. ☐ Free at both ends.

☐ Fixed at one end, free at the other.

Problem 9.

Five very long strings are carrying ***traveling*** waves from left to right. The first "reference" string 0 has mass density μ_0, tension T_0 and the wave it carries has wavelength λ_0 and amplitude A_0. The following table contains data about the waves on the other four strings, labelled a-d, in terms of the reference string.

String	μ	T	λ	ω	A	$\frac{E}{\lambda}$	P
0 (ref)	μ_0	T_0	λ_0		A_0		
a	$2\mu_0$	$2T_0$	λ_0		$\frac{1}{4}A_0$		
b	$\frac{1}{2}\mu_0$	$2T_0$	λ_0		A_0		
c	μ_0	$\frac{1}{2}T_0$	$2\lambda_0$		A_0		
d	$2\mu_0$	T_0	$\frac{1}{2}\lambda_0$		$2A_0$		

First, (in a copy of the table) ***fill in the missing information for the reference string***. Specifically, find ω_0, E_0/λ_0, and P_0 ***in terms of the reference string givens***.

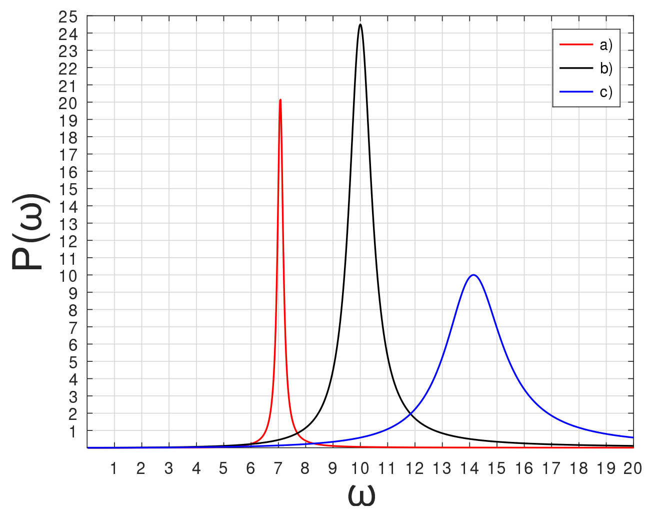
Second, fill in (a copy of) the table for the other four strings with answers formatted to be a pure dimensionless number times the value for that quantity for the reference string. For example, perhaps $\boldsymbol{\omega_d = 3.14 \times \omega_0}$ (but probably isn't).

Third, ***rank the average energy densities and power*** of strings a-d from least to greatest in (a copy of) the boxes provided below. Sort any that are equal by the value of the letter – if $\omega_c = \omega_b < \omega_a < \omega_d$ (unlikely) then fill in b,c,a,d in that order

Average $\frac{\textbf{Energy}}{\textbf{Length}}$:

Average Power:

Problem 10.



Resonance is an important concept in waves on a string, as *standing* waves have a set of discrete *natural frequencies* and can be driven at those frequencies (by e.g. the bow of a violin) or started in the shape of a mode at $t = 0$ (e.g. a piano string struck by a hidden hammer) and then vibrate, weakly damped, as they transfer energy into *sound* waves in string instruments!

For that reason, let's review Q and how it relates to the power curve as a function of frequency. Musical instrument strings are often engineered to have a relatively high Q so that they persist when a note is plucked, or struck, for a long time!

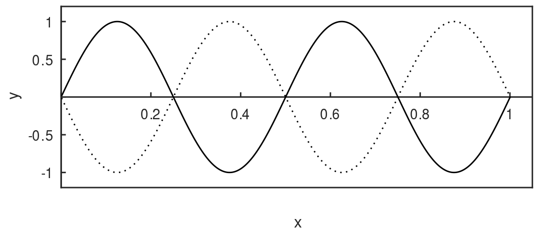
In the figure to the side three ***resonance curves*** are drawn showing the power $P_{a,b,c}(\omega)$ delivered to three distinct steady-state driven oscillators (or oscillation ***modes*** of a string): a) (red), b) (black), and c) (blue). Put down an *estimate* of the Q-value of each oscillator by looking at the graph. It may help for you to put down the definition of Q most relevant to the process of estimation on the page. Which string would produce music that died out slowly with the driving force is removed? Which string would produce music that dies out quickly?

$Q_a = \boxed{}$ dies out ☐ fastest ☐ slowest ☐ in between

$Q_b = \boxed{}$ dies out ☐ fastest ☐ slowest ☐ in between

$Q_c = \boxed{}$ dies out ☐ fastest ☐ slowest ☐ in between

Advanced Problem 11.



A string of total mass M and total length L is fixed at both ends, stretched so that the speed of waves on the string is v. It is vibrating in its $n = 4$ harmonic mode

$$y(x, t) = Asin(k_4 x)cos(\omega_4 t)$$

(as illustrated above on a scale where $A = 1$ and $L = 1$).

a) Following the book, ***derive*** the kinetic energy per wavelength, the potential energy per wavelength, and the total energy per wavelength as functions of time for arbitrary standing waves on a string fixed at both ends. The result shouldn't refer to the specific 4th harmonic; use ω_n and/or k_n – but should be specific to *standing waves* that satisfy *these boundary conditions*.

b) Use your result to find the total energy in the string for the $n = 4$th harmonic *specifically* in terms of M, L, T and A. Does it vary in time? How does it scale with these parameters? Does your answer have the right dimensions?

You may find it useful to remember that:

$$\int_0^{n\pi} \sin^2(u)du = \int_0^{n\pi} \cos^2(u)du = \frac{n\pi}{2}$$

Final Exam AND Second Hour Exam!

Continue studying for the final exam! Only ***two more weeks of class*** (and two chapters) to go in this textbook! Don't wait until the last minute to start!

Week 11: Sound

1.18: Sound Summary

- **Bulk Modulus**

  The ***Bulk Modulus*** of a fluid (or a solid) is defined by:

$$\Delta P = -B\frac{\Delta V}{V} \tag{11.1}$$

  where B is the bulk modulus. The bulk modulus of a solid is clearly related to Young's modulus, where compressive forces are applied in all three dimensions at once. In a gas, it tells us how much the gas compresses (at a given temperature) when the pressure in the gas is changed.

  For an ideal gas at fixed temperature, we can easily derive B using calculus:

$$P = \frac{NkT}{V} \tag{11.2}$$

  so

$$\frac{dP}{dV} = -\frac{NkT}{V^2} \tag{11.3}$$

  and

$$dP = -\frac{NkT}{V}\frac{dV}{V} = -P\frac{dV}{V} \tag{11.4}$$

  and we see that the bulk modulus of an ideal gas *at constant temperature* equals its pressure!

- **Speed of Sound** in a fluid

$$v = \sqrt{\frac{B}{\rho}} \tag{11.5}$$

  where B is the bulk modulus of the fluid and ρ is the density.

  In an ideal *monoatomic* gas

$$v = \sqrt{\frac{P}{\rho}} \tag{11.6}$$

  We usually assume that sound waves are ***adiabatic*** (energy conserving), not ***isothermal***. Also, liquids are nothing like ideal gases!

  For near-ideal-gases (such as air) we will use:

$$B = \gamma P \tag{11.7}$$

where $\gamma = 1.4$ for diatomic gases (see a chapter on thermodynamics for the derivation and explanation of γ). For liquids and non-ideal gases we will ignore further discussions of the thermodynamics and use B, or v itself, as a given.

- **Speed of Sound in Air**

  Air is to a decent approximation an ideal *diatomic* gas). $\gamma = 1.4$ for an ideal diatomic gas, so:

$$v = \sqrt{\frac{B}{\rho}} = \sqrt{\gamma \frac{P}{\rho}} \approx = \sqrt{1.4 \frac{10^5}{1.225}} \approx 343 \quad \text{m/sec} \tag{11.8}$$

  where we've used $P \approx 10^5$ pascals and $\rho = 1.225$ kg/m^3 for dry air at one atmosphere. This is in *very good agreement* with observation.

  We will often use approximations for this in class to facilitate arithmetic "in your head", such as $v_a \approx 340$ m/sec or $v_a \approx 333$ m/sec (so that $v_a \approx 1/3$ km/sec). This is very close to $v_a = 1000$ ft/sec in the English system, or $v_a = 1/5$ mi/sec, which can be quite useful for estimating distances to approaching thunderstorms.

- **Travelling Sound waves:**

  Plane (displacement) waves (in the x-direction):

$$s(x,t) = s_0 \sin(kx - \omega t)$$

  Spherical waves:

$$s(r,t) = s_0 \frac{R}{(\sqrt{4\pi})\, r} \sin(kr - \omega t)$$

  where R is a reference length needed to make the units right corresponding physically to the "size of the source" (e.g. the smallest ball that can be drawn that completely contains the source). The $\sqrt{4\pi}$ is needed so that the intensity has the right functional form for a spherical wave (see below).

- **Pressure Waves:** The pressure waves that correspond to these two displacement waves are:

$$P(x,t) = P_0 \cos(kx - \omega t)$$

  and

$$P(r,t) = P_0 \frac{R}{(\sqrt{4\pi})\, r} \cos(kr - \omega t)$$

  where $P_0 = v_a \rho \omega s_0 = Z \omega s_0$ with $Z = v_a \rho = \sqrt{B\rho}$ (a conversion factor that scales microscopic displacement to pressure). Note that:

$$P(x,t) = Z \frac{d}{dt} s(x,t)$$

  or, the displacement wave is a scaled derivative of the pressure wave.

  The pressure waves represent the oscillation of the pressure *around* the baseline ambient pressure P_a, e.g. 1 atmosphere. The total pressure is really $P_a + P(x,t)$ or $P_a + P(r,t)$ (and we could easily put a "Δ" in front of e.g. $P(r,t)$ to emphasize this point but don't so as to not confuse variation around a baseline with a derivative).

- **Sound Intensity:** The intensity of sound waves can be written:

$$I = \frac{1}{2} v_a \rho \omega^2 s_0^2$$

or

$$I = \frac{1}{2} \frac{1}{v_a \rho} P_0^2 = \frac{1}{2Z} P_0^2$$

- **Spherical Waves:** The intensity of spherical sound waves drops off like $1/r^2$ (as can be seen from the previous two points). It is usually convenient to express it in terms of the ***total power emitted by the source*** $\boldsymbol{P_{\mathrm{tot}}}$ as:

$$I = \frac{\boldsymbol{P_{\mathrm{tot}}}}{4\pi r^2}$$

This is "the total power emitted divided by the area of the sphere of radius r through which all the power must symmetrically pass" and hence it *makes sense!* One can, with some effort, take the intensity at some reference radius r and relate it to P_0 and to s_0, and one can easily relate it to the intensity at other radii.

- **Decibels:** Audible sound waves span some 20 orders of magnitude in intensity. Indeed, the ear is barely sensitive to a *doubling* of intensity – this is the smallest change that registers as a change in audible intensity. This motivate the use of ***sound intensity level*** measured in ***decibels***:

$$\beta = 10 \log_{10} \left(\frac{I}{I_0} \right) \ (\mathrm{dB})$$

where the reference intensity $I_0 = 10^{-12}$ watts/meter^2 is the ***threshold of hearing***, the weakest sound that is audible to a "normal" human ear.

Important reference intensities to keep in mind are:

  - 60 dB: Normal conversation at 1 m.
  - 85 dB: Intensity where long term continuous exposure *may* cause gradual hearing loss.
  - 120 dB: Hearing loss is ***likely*** for anything more than brief and highly intermittent exposures at this level.
  - 130 dB: Threshold of ***pain***. Pain is bad.
  - 140 dB: Hearing loss is ***immediate and certain*** – you are actively losing your hearing during any sort of prolonged exposure at this level and above.
  - 194.094 dB: The upper limit of undistorted sound (overpressure equal to one atmosphere). This loud a sound will instantly rupture human eardrums 50% of the time.

- **Doppler Shift: Moving Source**

$$f' = \frac{f_0}{(1 \mp \frac{v_s}{v_a})} \tag{11.9}$$

where f_0 is the unshifted frequency of the sound wave for receding (+) and approaching (-) source, where v_s is the speed of the source and v_a is the speed of sound in the medium (air).

- **Doppler Shift: Moving Receiver**

$$f' = f_0(1 \pm \frac{v_r}{v_a}) \tag{11.10}$$

where f_0 is the unshifted frequency of the sound wave for receding (-) and approaching (+) receiver, where v_r is the speed of the source and v_a is the speed of sound in the medium (air).

- **Stationary Harmonic Waves**

$$y(x,t) = y_0 \sin(kx) \cos(\omega t) \tag{11.11}$$

for displacement waves in a pipe of length L closed at one or both ends. This solution has a node at $x = 0$ (the closed end). The permitted resonant frequencies are determined by:

$$kL = n\pi \tag{11.12}$$

for $n = 1, 2...$ (both ends closed, nodes at both ends) or:

$$kL = \frac{2n-1}{2}\pi \tag{11.13}$$

for $n = 1, 2, ...$ (one end closed, nodes at the closed end).

- **Beats** If two sound waves of equal amplitude and slightly different frequency are added:

$$\begin{aligned} s(x,t) &= s_0 \sin(k_0 x - \omega_0 t) + s_0 \sin(k_1 x - \omega_1 t) && (11.14) \\ &= 2s_0 \sin(\frac{k_0 + k_1}{2} x - \frac{\omega_0 + \omega_1}{2} t) \cos(\frac{k_0 - k_1}{2} x - \frac{\omega_0 - \omega_1}{2} t) && (11.15) \end{aligned}$$

which describes a wave with the average frequency and twice the amplitude modulated so that it "beats" (goes to zero) at the difference of the frequencies $\delta f = |f_1 - f_0|$.

11.1: Sound Waves in a Fluid

Waves propagate in a fluid much in the same way that a disturbance propagates down a closed hall crowded with people. If one shoves a person so that they knock into their neighbor, the neighbor falls against *their* neighbor (and shoves back), and their neighbor shoves against their still further neighbor and so on.

Such a wave differs from the transverse waves we studied on a string in that the displacement of the medium (the air molecules) is *in the same direction* as the direction of propagation of the wave. This kind of wave is called a *longitudinal* wave.

Although different, sound waves can be related to waves on a string in many ways. Most of the similarities and differences can be traced to one thing: a string is a one dimensional medium and is characterized only by length; a fluid is typically a three dimensional medium and is characterized by a volume.

Air (a typical fluid that supports sound waves) does not support "tension", it is under pressure. When air is compressed its molecules are shoved closer together, altering its density and occupied volume. For small changes in volume the pressure alters approximately *linearly* with a coefficient called the “bulk modulus” B describing the way the pressure increases as the fractional volume decreases. Air does not have a mass per unit length μ, rather it has a mass per unit volume, ρ.

The velocity of waves in air (treated as an ideal diatomic gas at 1 atmosphere) is given by:

$$v = \sqrt{\frac{B}{\rho}} = \sqrt{\gamma\frac{P}{\rho}} \approx= \sqrt{1.4\frac{10^5}{1.225}} \approx 343 \quad \mathrm{m/sec} \tag{11.16}$$

The “approximately” here is fairly serious. The speed obviously varies like the square root of the air pressure over the density, which both vary significantly with altitude and with the weather at any given altitude as low and high pressure areas move around on the earth’s surface. Both also vary with the temperature (hotter molecules push each other apart more strongly at any given density). Consequently the speed of sound can vary by a few percent from the approximate value given above over the course of as little as a day at any given location, in addition to varying by more than that as one travels from sea level to the tops of mountains.

11.2: The Wave Equation for Sound

The derivation of the wave equation for sound in a gas is, to put it bluntly, “difficult”. It involves synthesizing three separate ideas – the “state equation” for the gas (and the concept of an “adiabatic process” expressed in the calculus of a bulk medium), the law of conservation of mass (the continuity equation), and a force law – Newton’s second law for a chunk of the fluid. If you are a non-physics major student (reading this) for whom things like this are moderately terrifying, be at peace – as was the case for the derivation of the solutions to the harmonic oscillator equation above, you will not be held responsible in any way for deriving the wave equation for sound, and you can without any penalty skip from the break just below to the similar line ending the break below. Even though you won’t be held responsible for this, though, you may find it interesting. Physics majors absolutely should go through the derivation and try to understand it, and it is up to their instructor as to whether or not it is required.

The Following is an Advanced Topic and May be Skipped!

We start with the physicist version of the Ideal Gas Law, as the equation of state of an ideal gas. Air at standard temperature and pressure is a very good approximation of an ideal gas, as are most gases far away from the liquid-gas phase transition

$$PV = NkT \tag{11.17}$$

We assume that this equation applies to any small chunk ΔV of the gas large enough to contain many molecules but small enough to be (eventually) treatable as a differential volume. You will recognize this as the coarse-graining hypothesis that we’ve used from the beginning to describe microscopically discrete matter as ”mass density”.

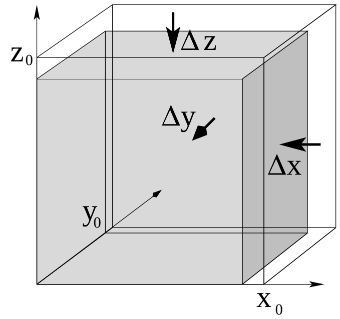


Figure 11.1: A volume $V = xyz$ containing N molecules is compressed across all of its faces so that it shrinks by Δx in the x-direction, Δy in the y-direction, and Δz in the z-direction. This reduces its volume by $\Delta V = xy\Delta z + yz\Delta x + zx\Delta y$.

Next, we consider what happens when we compress this volume of gas. Suppose that a fixed number N of gas molecules occupy a box with sides x, y, and z so that its volume $V_0 = x_0 y_0 z_0$. Such a box is pictured in figure 11.1. Now imagine that the box is compressed (by pressure in the surrounding fluid) so that it shrinks by a (small!) Δx in the x-direction, Δy in the y-direction, and Δz in the z-direction while the number of molecules in the box does not change. The change in volume of the box is $\Delta V = x_0 y_0 z_0 - (x_0 - \Delta x)(y_0 - \Delta y)(z_0 - \Delta z) \approx x_0 y_0 \Delta z + y_0 z_0 \Delta x + z_0 x_0 \Delta y$ plus terms with products of small pieces such as $x_0 \Delta y \Delta z$ which we will neglect.

Extending our previous discussion of Young's Modulus, stress, and strain to three dimensions, we write (for the three directions independently):

$$\begin{aligned}
\Delta P &= \frac{\Delta F_x}{y_0 z_0} = -Y\frac{\Delta x}{x_0} \\
\Delta P &= -Y\frac{\Delta y}{y_0} \\
\Delta P &= -Y\frac{\Delta z}{z_0}
\end{aligned}$$

If we multiply each term by a clever form of 1:

$$\begin{aligned}
\Delta P &= -Y\frac{\Delta x y_0 z_0}{x_0 y_0 z_0} \\
\Delta P &= -Y\frac{\Delta y x_0 z_0}{x_0 y_0 z_0} \\
\Delta P &= -Y\frac{\Delta z x_0 y_0}{x_0 y_0 z_0}
\end{aligned}$$

and then add them and divide both sides by 3, we get:

$$\Delta P = \frac{-Y}{3}\frac{\Delta V}{V_0} = -B\frac{\Delta V}{V_0} \tag{11.18}$$

which relates a new quantity, the ***bulk modulus*** B, to Young's modulus. This equation holds for solids, liquids and gases, but we only need to use it for our ideal gas above. This is one of the key equations needed to derive the wave equation for sound. Note well that the bulk modulus has units of pressure in pascals (or atmospheres or bar or torr), just as do Young's modulus and the shear modulus.

In this equation *for* an ideal gas, $\Delta P = P - P_0$, the increase in pressure relative to a "reference pressure" P_0 associated with the number of molecules N, the "reference volume" volume V_0 (which is really any small chunk of volume in the bulk gas, it needn't be an actual box of gas with solid walls), and temperature T of the gas:

$$P_0 V_0 = NkT \tag{11.19}$$

If we divide the volume over to the other side and multiply both sides by the molecular mass m we obtain the (reference) density of the gas at the reference pressure/volume:

$$\rho_0 = \frac{Nm}{V_0} = P_0 \frac{kT}{m} \tag{11.20}$$

Let's linearize the change in density in terms of the change in volume using the binomial expansion, holding (in our minds) the number of molecules in the volume unchanged. "Linearizing" means that we just keep the first term in ΔV (the linear term) and ignore the terms that are higher order in ΔV^2, etc:

$$\begin{aligned}
\Delta\rho &= \rho - \rho_0 = \frac{mN}{V_0 - \Delta V} - \frac{mn}{V_0} \\
&= \frac{mN}{V_0}(1 - \frac{\Delta V}{V_0})^{-1} - \frac{mn}{V_0} = \frac{mN}{V_0}\left(1 + \frac{\Delta V}{V_0} + \frac{\Delta V^2}{2V_0^2} + ...\right) - \frac{mn}{V_0} \\
&= \frac{mN}{V_0}\frac{\Delta V}{V_0} + ... \\
&\approx \rho_0 \frac{\Delta V}{V_0}
\end{aligned} \tag{11.21}$$

Note well that ΔV is the *positive* magnitude of the *reduction* in volume according to our use of signs. This means that :

$$\frac{\Delta\rho}{\rho_0} = \frac{\Delta V}{V_0}$$

As one might expect, reducing the volume increases the density in identical ways relative to the initial values. With the same sign convention for ΔV, we also have:

$$\Delta P = B\frac{\Delta V}{V_0} = B\frac{\Delta\rho}{\rho_0} \tag{11.22}$$

where reducing the volume increases the density and increases the pressure of the fluid. This equation *also* holds for bulk compression of ANY substance, solid, gas or fluid.

The definition of B and its extension to ρ is not complete or universal. For example, it does not account for temperature changes, which can also cause pressure changes at a constant volume! Volume changes can also occur at a constant pressure if one varies the temperature. There are literally an infinite number of ways one can compress a chunk of gas (doing work) and distribute the work as a mix of an increase in the average kinetic energy (temperature) of molecules in the gas and "heat", energy that flows in or out through the sides of the volume. The two limiting cases of this are ***isothermal*** compression where *all* of the work done to compress the fluid flows out of the volume as heat (keeping the volume at a constant temperature) and ***adiabatic*** (or "isentropic") compression where *none* of the work goes out of the volume; it remains as an increase in average kinetic energy, hence increases the absolute temperature.

Sound waves are not precisely either of these "pure" processes, but dry air at one atmosphere is a *good thermal insulator* and most sound waves vary on timescales that are short relative to the time required for significant heat to flow. We therefore assume that sound waves in air are composed of ***adiabatic compression*** of small volumes of the air (treated as an ideal gas) by the surrounding air.

We pause for a very brief discussion of adiabatic processes for a volume of ideal gas with a constant number of particles. The justification for the starting equation is beyond the scope of this course but is to be found in any introductory physics textbook that covers the thermodynamics of an ideal gas. The relation between pressure and volume in an adiabatic compression of an ideal gas is:

$$PV^{\gamma} = A \tag{11.23}$$

where $\gamma = 5/3 = 1.67$ for monoatomic ideal gases and $\gamma = 1.4$ for diatomic ideal gases (such as air) and A is a constant with the appropriate units. If we form the differential of both sides, we get:

$$\gamma P V^{\gamma-1} dV - V^{\gamma} dP = 0 \tag{11.24}$$

or (dividing both sides by V^{γ} and rearranging):

$$dP = \gamma P \frac{dV}{V} \tag{11.25}$$

or (comparing with the equations defining the bulk modulus above):

$$B = \gamma P \tag{11.26}$$

This will make it easy to at least estimate the expected speed of sound in air at the end.

Let's return to the expression defining the bulk modulus above, evaluating the (small) changes ***relative to*** the reference pressure and density for air, P_0 and ρ_0 (eventually one atmosphere and 1.225 kg/m^3 at room temperature and sea level).

$$\Delta P = P - P_0 = B\frac{(\rho - \rho_0)}{\rho_0} = B\frac{\Delta\rho}{\rho_0} \tag{11.27}$$

We now make a simplifying change of variables. Sound waves are ultimately going to be described by a change in pressure *relative* to some reference/background pressure and changes in density *relative* to the corresponding reference/background density. Carrying along the reference values and/or constantly writing Δ's will be tedious. So we now define the ***relative pressure change***:

$$p = P - P_0 \tag{11.28}$$

which I will sometimes refer to as the "overpressure", although it can be positive or negative and hence may be an underpressure as well. We will also refer to the (dimensionless!) ***relative density change*** scaled by the reference density, which is also called the "condensation" (hence symbol c):

$$c = \frac{\rho - \rho_0}{\rho_0} \tag{11.29}$$

Thus

$$p = Bc \tag{11.30}$$

or in words, the overpressure is the bulk modulus times the condensation.

Next, we need a fundamental concept from our discussion of fluid flow. Suppose we have a small block of fluid:

$$\Delta V = \Delta x \Delta y \Delta z$$

and imagine holding its *volume constant*. The only way the density of the fluid in the block can change if the volume is held constant is by fluid particles *flowing into or out of the block*. That is, let us think for a moment about what happens when we do *not* insist that the number of molecules in a small block of fluid is constant.

We are only concerned at the moment with a single dimension – the direction of the ***longitudinal*** motion of molecules in the sound wave – so we can consider only the motion of particles in the x direction and assume that at least, $\bar{v}_y = \bar{v}_z = 0$, there is no net motion in the y and z directions associated with the sound wave. Conservation of mass now requires that the change in the number of particles *in* the volume ΔV equals the number flowing in on the left face minus the number flowing out on the right face. Subject to these conditions this works out to become the ***continuity equation***:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} = 0 \tag{11.31}$$

where $u = \bar{v}_x$ is the local average x-directed velocity of the fluid. More generally (allowing for nonzero average motion in the other two directions) one obtains the three dimensional form:

$$\frac{\partial \rho}{\partial t} + \vec{\boldsymbol{\nabla}} \cdot \rho \vec{\boldsymbol{u}} = 0 \tag{11.32}$$

but we won't need this in this simplified treatment. This is basically the law of conservation of mass applied to the fluid.

Again we linearize this so that we express the changes in density relative to the reference density, using the condensation c as defined above to substitute $\rho = \rho_0 + \rho_0 c$ into both terms:

$$\frac{\partial}{\partial t}(\rho_0 + \rho_0 c) + \frac{\partial}{\partial x}(\rho_0 u + \rho_0 c u) = 0 \tag{11.33}$$

Note that ρ_0 is a *constant* as far as these derivatives are concerned. We can therefore cancel it out of the entire expression and rewrite it relating only c and u:

$$\frac{\partial c}{\partial t} + \frac{\partial}{\partial x}(u + cu) = 0 \tag{11.34}$$

c is necessarily "small" for the linearization above to work, so the product $cu \ll u$. We neglect this term and end up with:

$$\frac{\partial c}{\partial t} + \frac{\partial u}{\partial x} = 0 \tag{11.35}$$

or, the time derivative of the relative change in density is the negative space derivative of the (x) velocity of the fluid at each point. This makes sense.

Next, we need what amounts to "Newton's Second Law" for a chunk of fluid. Suppose again we imagine a chunk of fluid with sides $\Delta x, \Delta y, \Delta z$. The force on this chunk in the x-direction is basically:

$$F_x = P(x)\Delta y \Delta z - P(x + \Delta x)\Delta y \Delta z = -\frac{\partial P}{\partial x}\Delta x \Delta y \Delta z \tag{11.36}$$

Newton's Law becomes:

$$F_x = -\frac{\partial P}{\partial x}\Delta V = \frac{d(\Delta m u)}{dt} = \rho \Delta V \frac{du}{dt} \tag{11.37}$$

where as before, u is the x-component of the local velocity only. This is called the Euler force equation. We cancel the ΔV and rearrange this into:

$$\rho \frac{du}{dt} + \frac{\partial P}{\partial x} = 0 \tag{11.38}$$

Again, we have to linearize this around ρ_0 and use the relevant part of the *total derivative with respect to* t (this is multivariate calculus you may not have learned yet):

$$(\rho_0 + \rho_0 c)(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}) + \frac{\partial (P_0 + p)}{\partial x} = 0 \tag{11.39}$$

In this equation (recall) c is very small. P_0 is the background pressure, presumed constant. The variation of u with x is small. Neglecting or cancelling these terms gives us:

$$\rho_0 \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = 0 \tag{11.40}$$

which we can view as the linearized Euler equation, a.k.a. Newton's second law for a coarse grained chunk of the fluid small enough to be treated as a differential, but large enough that densities and pressures make sense.

If we line up these two equations:

$$\begin{aligned} \frac{\partial c}{\partial t} + \frac{\partial u}{\partial x} &= 0 \\ \rho_0 \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} &= 0 \end{aligned}$$

and take the partial time derivative of the first and the partial space derivative of the second, we get:

$$\begin{aligned} \frac{\partial^2 c}{\partial t^2} + \frac{\partial^2 u}{\partial t \partial x} &= 0 \\ \rho_0 \frac{\partial^2 u}{\partial t \partial x} + \frac{\partial^2 p}{\partial x^2} &= 0 \end{aligned}$$

If we divide the second equation by ρ_0 and subtract, we get:

$$\frac{\partial^2 c}{\partial t^2} - \frac{1}{\rho}\frac{\partial^2 p}{\partial x^2} = 0 \tag{11.41}$$

Finally, we multiply both sides by B and use our linearized state equation $p = Bc$ to eliminate c (the condensation) in favor of p:

$$\frac{\partial^2 p}{\partial t^2} - \frac{B}{\rho}\frac{\partial^2 p}{\partial x^2} = \frac{\partial^2 p}{\partial t^2} - v^2 \frac{\partial^2 p}{\partial x^2} = 0 \tag{11.42}$$

or (rearranging to a more or less standard form):

$$\frac{\partial^2 p}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 p}{\partial t^2} = 0 \tag{11.43}$$

with:

$$v = \sqrt{\frac{B}{\rho_0}} \tag{11.44}$$

Note that this contains an implicit temperature dependence since $B = \gamma P = \gamma NkT/V$. Thus:

$$v = \sqrt{\frac{B}{\rho_0}} = \sqrt{\frac{\gamma kT}{M/N}} = \sqrt{\frac{\gamma kT}{m_{\rm mol}}}$$

where $m_{\rm mol}$ is the molecular mass (kg/molecule). Alternatively, one can scale by Avogadro's number to obtain the same expression in terms of molar mass and the ideal gas constant R:

$$v = \sqrt{\frac{\gamma kT}{m_{\rm mol}}} = \sqrt{\frac{\gamma RT}{M_{\rm mol}}} \tag{11.45}$$

where $M_{\rm mol}$ is in kg/mole. For dry air in the adiabatic approximation, this works out to be:

$$v = 20.05\sqrt{T} \tag{11.46}$$

(with T given in degrees kelvin).

Finally, since we can express sound as *either* a pressure wave *or* as a displacement wave, we have to find a relationship between s (displacement of some individual arbitrary (ith) molecule from a linearized "equilibrium position" s_i) and p. We do this using $u = \frac{\partial s}{\partial t}$ in the x-direction and:

$$\begin{aligned}
\rho_0 \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} &= 0 \\
\rho_0 \frac{\partial^2 s}{\partial t^2} + \frac{\partial p}{\partial x} &= 0 \\
\rho_0 \frac{\partial^3 s}{t^2 x} + \frac{\partial^2 p}{\partial x^2} &= 0 \\
\rho_0 \frac{\partial^3 s}{t^2 x} + \frac{1}{v^2}\frac{\partial^2 p}{\partial t^2} &= 0 \\
\frac{\partial^2 (\rho_0 \frac{\partial s}{\partial x} + \frac{1}{v^2} p)}{\partial t^2} &= 0
\end{aligned}$$

There are infinitely many solutions for s from this, but they boil down to the freedom to make s_i any constant position in the fluid and to add any constant speed u_0 to the entire fluid – the usual freedom for zero-net-force problems. The condition we are interested (which is sufficient, though not unique):

$$\rho_0 \frac{\partial s}{\partial x} = -\frac{1}{v^2} p \tag{11.47}$$

Suppose

$$p(x,t) = p_0 \cos(kx - \omega t) \tag{11.48}$$

(a simple harmonic travelling wave). Then:

$$\begin{aligned}
s(x,t) &= \frac{1}{v^2 \rho_0} p_0 \int \cos(kx - \omega t) dx \\
s(x,t) &= \frac{1}{kv^2 \rho_0} p_0 \sin(kx - \omega t) = \frac{1}{\omega v \rho_0} p_0 \sin(kx - \omega t) \\
s(x,t) &= \frac{1}{Z\omega} p_0 \sin(kx - \omega t) = s_0 \sin(kx - \omega t)
\end{aligned}$$

(plus, as noted above, any $u_0 t + s_i$ term you like) where:

$$Z = \rho_0 v \tag{11.49}$$

and:

$$p_0 = Z s_0 \tag{11.50}$$

Obviously, displacement will satisfy an identical wave equation to pressure, but the pressure and displacement solutions are $\pi/2$ out of phase – where displacement is $s_0 \sin(kx - \omega t)$, pressure is $p_0 \cos(kx - \omega t) = p_0 \sin(kx - \omega t + \pi/2$.

There are a number of ways to write the relationship between p and s as derivatives. We already have:

$$p(x,t) = -v^2 \rho_0 \frac{\partial s}{\partial x} = -Zv \frac{\partial s}{\partial x} \tag{11.51}$$

We can also take a partial with respect to time of the $s(x,t)$ result we obtained to get:

$$\frac{\partial s}{\partial t} = \frac{1}{Z\omega} p_0 \frac{\partial \sin(kx - \omega t)}{\partial t} = -\frac{1}{Z} p_0 \cos(kx - \omega t) \tag{11.52}$$

or:

$$p(x,t) = Z \frac{\partial s}{\partial t} \tag{11.53}$$

Note Well: $s(x,t)$ is the ***relative*** displacement of any given molecule in the fluid from an equilibrium position s_i. s_i itself is a function of x. s_0 is the amplitude of the oscillation of molecules at the equilibrium position of x from this position. All of this makes no sense unless s_0 is small, its mean velocity is small, and hence x can accurately pick out a small block of fluid in the coarse grained linearized limit! The pressure wave is a little bit easier to understand than the displacement wave, as bulk displacement can occur even in a fluid with no waves – e.g. wind.

It is beyond the scope of this course to go any further into this, which is (as you can see) a derivation where we repeatedly had to throw away large terms. A similar derivation (essentially the derivation above repeated in all three dimensions) can be used to obtain a three-dimensional wave equation:

$$\nabla^2 p - \frac{1}{v^2} \frac{\partial^2 p}{\partial t^2} = 0$$

that has as solutions (among others) the symmetric spherical waves indicated below. Beyond this, the next stop in fluid dynamics is the full Navier-Stokes equation, which is so difficult that it cannot be generally solved (yet), where the results above are basically the results of linearizing it and neglecting certain things.

If You Skipped The Section Above, Start Reading Again Here!

Let's summarize the from the (omitted) section above:

- For a longitudinal plane wave, the pressure at a longitudinal point x at time t is given by solutions to the one dimensional wave equation:

$$\frac{\partial^2 p}{\partial x^2} - \frac{1}{v^2} \frac{\partial^2 p}{\partial t^2} = 0 \tag{11.54}$$

with:

$$v = \sqrt{\frac{B}{\rho_0}} = \sqrt{\frac{\gamma k T}{m_{\rm mol}}} = \sqrt{\frac{\gamma R T}{M_{\rm mol}}} \tag{11.55}$$

where $\gamma = 7/5$ for a diatomic gas and 5/3 for a monoatomic gas, $m_{\rm mol}$ is in kg/molecule, $M_{\rm mol}$ is in kg/mole, k is Boltzmann's constant, $R = N_A k$ is the ideal gas constant, and T is in degrees kelvin/absolute.

For dry air in the adiabatic approximation, this works out to be:

$$v = 20.05\sqrt{T} = 343 \text{ m/sec at } T = 20^\circ \text{ K} \tag{11.56}$$

- We can find the *relative* displacement of molecules whose equilibrium position is x at time t from the pressure or vice versa from the relations:

$$p(x,t) = -v^2\rho_0 \frac{\partial s}{\partial x} = -Zv\frac{\partial s}{\partial x} \tag{11.57}$$

or

$$\frac{\partial s}{\partial t} = -\frac{1}{Z} p_0 \cos(kx - \omega t) \tag{11.58}$$

or:

$$p(x,t) = Z\frac{\partial s}{\partial t} \tag{11.59}$$

where

$$Z = \rho_0 v = \sqrt{\rho_0 B} = \sqrt{\gamma \rho_0 P_0} \tag{11.60}$$

(the latter for an ideal gas in the adiabatic limit only). Pressure waves and displacement waves are thus ***out of phase by*** $\pi/2$ with the pressure wave ***leading*** the displacement wave by this amount.

- If we manipulate the last two equations, it is easy to show that displacement ***also*** satisfies the *same* wave equation with the *same* speed $v = \sqrt{B/\rho_0}$:

$$\frac{\partial^2 s}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 s}{\partial t^2} = 0 \tag{11.61}$$

11.3: Sound Wave Solutions

From the previous section (whether or not you read it) sound waves can be characterized one of two ways: as organized fluctuations in the *displacement* of the molecules of the fluid as they oscillate around an equilibrium position or as organized fluctuations in the *pressure* (or density/concentration, see above) of the fluid as molecules are crammed closer together or are diven farther apart than they are on average in the quiescent fluid. We can visualize this and understand it from figure 11.2 below.

One dimensional sound waves propagate in one direction (out of three) at any given point in space. This means that in the direction *perpendicular* to propagation, the wave is spread out to form a "wave front" – a planar region where all of the molecules are moving together.

In three dimensions, sounds waves satisfy a three-dimensional wave equation. In this case the wave front can be nearly arbitrary in shape initially (corresponding to the shape of

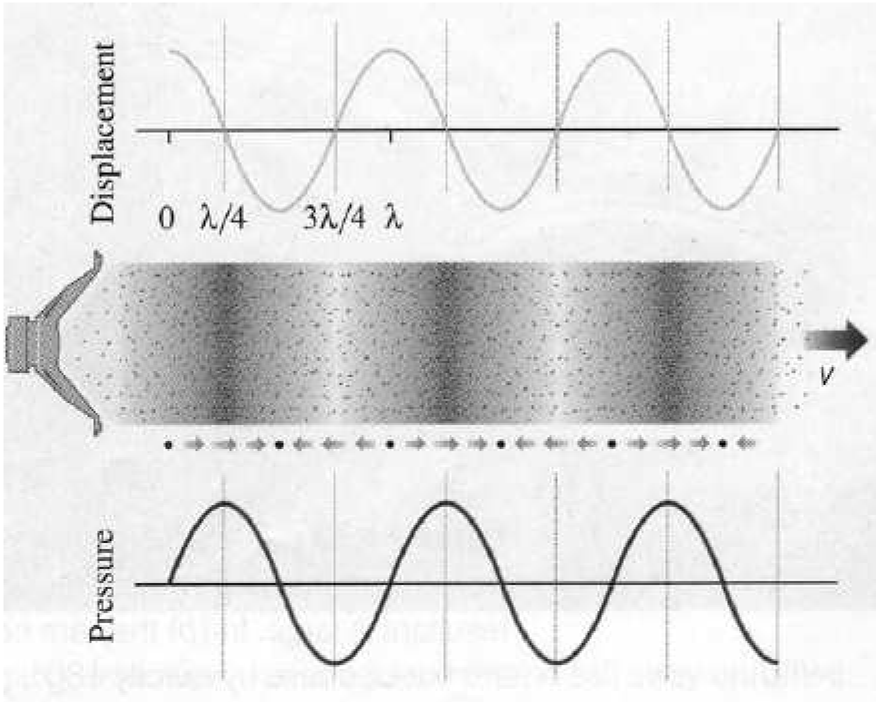


Figure 11.2: Schematic representation of a sound wave, illustrating how air molecules move ***longitudinally*** back and forth around their equilibrium position at the same time they generate ***over and under pressure waves*** around the background pressure of P_0 ~1 atm.

a speaker surface producing the wave, for example). Thereafter it evolves according to the mathematics of the wave equation in three dimensions, which is similar to but a wee bit more complicated than the wave equation in one dimension. Or possibly even a *lot* more complicated... more complicated than is appropriate for this course. However, we will indicate what one of the *simplest* solutions in three dimensions is like, the one appropriate to a "spherical speaker", or a sound source that radiates sound in all directions uniformly.

We don't need to do any more work to write these solutions down. All the hard part was done in the last chapter. For one-dimensional plane wave solutions we can make wave pulses or harmonic waves exactly like we did before, but now the "wave" is (over)pressure p or displacement s. For three I'll just give the appropriate solution for you to use below.

Let's consider ***only harmonic waves*** with a fixed frequency and wavelength, since wave pulses are not that common. Harmonic waves dominate our utilization of sound for communication and musical entertainment and awareness of our environment.

a) *Plane Wave* solutions. In these solutions, the entire wave moves in one direction (say the x direction) and the wave front is a 2-D plane perpendicular to the direction of propagation. These (displacement) solutions can be written as (e.g.):

$$s(x, t) = s_0 \sin(kx - \omega t) \tag{11.62}$$

where s_0 is the maximum displacement in the travelling wave (which moves in the x direction) and where all molecules in the entire *plane* at position x are displaced by the same amount.

Even spherical waves or waves produced by an arbitrary source behave like and are well described by plane waves near any given point in space. So are waves propagating down a constrained environment such as a tube that permits waves to only travel in "one direction". Most of the properties of sound of interest can be easily understood in terms of plane waves alone.

As we see above, we can *equally* well describe this specific sound plane wave in terms of its *pressure*, instead of *displacement*:

$$p(x,t) = P_{\rm atm} + p_0 \cos(kx - \omega t) \tag{11.63}$$

Now the wave amplitude p_0 describes the maximum ***overpressure*** of a wave oscillating around the equilibrium pressure of e.g. 1 atmosphere. Note well that pressure waves are $\boldsymbol{\pi/2}$ ***ahead of*** the displacement waves as I indicated by using cosine instead of sine. The pressure waves are at maxima and minima where the displacement is zero and vice versa.

b) *Spherical Wave* solutions. Sound is often emitted from a source that is highly localized (such as a hammer hitting a nail, or a loudspeaker). If the sound is emitted equally in all directions from the source, a spherical wavefront is formed. Even if it is not emitted equally in all directions, sound from a localized source will generally form a spherically curved wavefront as it travels away from the point with constant speed. The displacement of a spherical wavefront *decreases* as one moves further away from the source because the *energy* in the wavefront is spread out on larger and larger surfaces. Its form is given by:

$$s(r,t) = s_0 \frac{R}{r} \sin(kr - \omega t) \tag{11.64}$$

where R is a "reference length". In practical terms, R might be the radius of a spherical speaker's surface, and s_0 might be the amplitude of that surface's oscillation to create the wave. Once the wave leaves the surface, its amplitude (as we will see) has to diminish like $1/r$ in order for energy to be conserved.

We can also express this as a pressure wave.

$$p(r,t) = p_0 \frac{R}{r} \cos(kr - \omega t) \tag{11.65}$$

where r is the radial distance away from the point-like source. Note again that the pressure wave is $\pi/2$ out of phase with (ahead of) the displacement wave.

Remember that this pressure wave is oscillating *around* one atmosphere of pressure, so that the actual total air pressure is:

$$p_{\rm tot}(r,t) = P_{\rm atm} + p_0 \frac{R}{r} \cos(kr - \omega t) \tag{11.66}$$

just as the displacement wave is for the displacement of *each* molecule in the air relative to its equilibrium position.

The key thing to remember about both of these spherical wave solutions is that the pressure wave or displacement wave *amplitudes* die down like $\boldsymbol{1/r}$, where r is the distance of the observer from the source. Plane waves, on the other hand, do not change their amplitude as they propagate (or rather, they do so only very gradually due to damping phenomena in the air that removes energy from the wave to slightly heat the air).

Next, let's discuss the *energy* carried by sound waves, as it is very important to anyone who wants to talk or listen or make music.

11.4: Sound Wave Intensity

The energy density of sound waves is given by:

$$\frac{dE}{dV} = \frac{1}{2}\rho\omega^2 s^2 \tag{11.67}$$

(again, very similar in form to the energy density of a wave on a string). However, this energy per unit *volume* is propagated in a single direction. It is therefore spread out so that it crosses an *area*, not a single point. Just how much energy an object receives therefore depends on how much *area* it intersects in the incoming sound wave, not just on the energy density of the sound wave itself.

For this reason the energy carried by sound waves is best measured by *intensity*: the energy per unit time per unit area perpendicular to the direction of wave propagation. Imagine a box with sides given by ΔA (perpendicular to the direction of the wave's propagation) and $v\Delta t$ (in the direction of the wave's propagation. All the energy in this box crosses through ΔA in time Δt. That is:

$$\Delta E = (\frac{1}{2}\rho\omega^2 s^2)\Delta A v \Delta t \tag{11.68}$$

or

$$I = \frac{\Delta E}{\Delta A \Delta t} = \frac{1}{2}\rho\omega^2 s^2 v \tag{11.69}$$

which looks very much like the *power* carried by a wave on a string. In the case of a plane wave propagating down a narrow tube, it is very similar – the power of the wave is the intensity times the tube's cross section.

However, consider a spherical wave. For a spherical wave, the intensity looks something like:

$$I(r,t) = \frac{1}{2}\rho\omega^2 \frac{s_0^2(4\pi)R^2\sin^2(kr-\omega t)}{4\pi r^2} v \tag{11.70}$$

which can be written as:

$$I(r,t) = \frac{\boldsymbol{P}_{\mathbf{tot}}}{4\pi r^2} \tag{11.71}$$

where $\boldsymbol{P}_{\mathbf{tot}}$ is the total power in the wave. Expressing it this way helps a lot with all of those constants (the R and 4π and ρ etc.). All we really need to know is the total power emitted by the source in watts, and we can predict how the sound intensity will drop off with distance!

This makes sense from the point of view of energy conservation and symmetry. If a source emits a power $\boldsymbol{P}_{\mathbf{tot}}$, that energy has to cross each successive spherical surface that surrounds the source. Those surfaces have an area equal to $A = 4\pi r^2$. Thus the surface at $r = 2r_0$ has 4 times the area of one at $r = r_0$, but the *same* total power has to go through both surfaces. Consequently, the intensity at the $r = 2r_0$ surface has to be $1/4$ the intensity at the $r = r_0$ surface.

It is important to remember this argument, simple as it is. Next week we will learn about Newton's law of gravitation. There we will learn that the gravitational field diminishes as $1/r^2$ with the distance from the source. Electrostatic field also diminishes as $1/r^2$. There seems to be a shared connection between symmetric propagation and spherical geometry; this will form the basis for *Gauss's Law* in electrostatics and much beautiful math, as all of these ideas are *connected* by the geometry of the sphere.

11.4.1: Sound Displacement and Intensity In Terms of Pressure

The pressure in a sound wave (as noted) oscillates around the mean/baseline ambient pressure of the air (or water, or whatever). The pressure *wave* in sound is thus the time varying pressure *difference* – the amplitude of the pressure oscillation around the *mean* of normal atmospheric pressure.

As always, we will be interested in writing the pressure *as* a harmonic wave (where we can) and hence will use the peak pressure difference as the amplitude of the wave. We will call this pressure the (peak) "overpressure":

$$P_0 = P_{\max} - P_a \tag{11.72}$$

where P_a is the baseline atmospheric pressure. It is easy enough to express the pressure wave in terms of the displacement wave (and vice versa). The amplitudes are related by:

$$P_0 = v_a \rho \omega s_0 = Z \omega s_0 \tag{11.73}$$

where $Z = v_a \rho$, the product of the speed of sound in air and the air density. The pressure and displacement waves are $\boldsymbol{\pi/2}$ ***out of phase!*** with the pressure wave leading the displacement wave:

$$P(t) = P_0 \cos(kx - \omega t) \tag{11.74}$$

(for a one-dimensional "plane wave", use kr and put it over $1/r$ to make a spherical wave as before). The displacement wave is (Z times) the time derivative of the pressure wave, note well.

The intensity of a sound wave can also be expressed in terms of ***pressure*** (rather than displacement). The expression for the intensity is then very simple (although not so simple to derive):

$$I = \frac{P_0^2}{2Z} = \frac{P_0^2}{2 v_a \rho} \tag{11.75}$$

11.4.2: Sound Pressure and Decibels

Source of Sound	P (Pa)	I	dB
Auditory threshold at 1 kHz	2×10^{-5}	10^{-12}	0
Light leaf rustling, calm breathing	6.32×10^{-5}	10^{-11}	10
Very calm room	3.56×10^{-4}	3.16×10^{-12}	25
A Whisper	2×10^{-3}	10^{-8}	40
Washing machine, dish washer	6.32×10^{-3}	10^{-7}	50
Normal conversation at 1 m	2×10^{-2}	10^{-6}	60
Normal (Ambient) Sound	6.32×10^{-2}	10^{-5}	70
"Loud" Passenger Car at 10 m	2×10^{-1}	10^{-4}	80
Hearing Damage *Possible*	0.356	3.16×10^{-4}	85
Traffic On Busy Roadway at 10 m	0.356	3.16×10^{-4}	85
Jack Hammer at 1 m	2	10^{-2}	100
Normal Stereo at Max Volume	2	10^{-2}	100
Jet Engine at 100 m	6.32-200	10^{-1} to 10^{3}	110-140
Hearing Damage *Likely*	20	1	120
Vuvuzela Horn	20	1	120
Rock Concert ("The Who" 1982) at 32 m	20	1	120
Threshold of *Pain*	63.2	10	130
Marching Band (100-200 members, in front)	63.2	10	130
"Very Loud" Car Stereo	112	31.6	135
Hearing Damage *Immediate, Certain*	200	10^{2}	140
Jet Engine at 30 m	632	10^{3}	150
Rock Concert ("The Who" 1982) at *speakers*	632	10^{3}	150
30-06 Rifle 1 m to side	6,320	10^{5}	170
Stun Grenades	20,000	10^{6}	180
Limit of Undistorted Sound (human eardrums rupture 50% of time)	101,325 (1 atm)	2.51×10^{7}	194.094

Table 7: Table of (approximate) P_0 and sound pressure levels in decibels relative to the threshold of human hearing at 10^{-12} watts/m^2. Note that ordinary sounds only extend to a peak overpressure of $P_0 = 1$ atmosphere, as one cannot oscillate symmetrically to underpressures pressures *less* than a vacuum.

The one real problem with the very simple description of sound intensity given above is one of scale. The human ear is ***routinely exposed to and sensitive to*** sounds that vary by ***twenty orders of magnitude*** – from sounds so faint that they barely can move our eardrums to sounds so very loud that they immediately *rupture them!* Even this isn't the full range of sounds out there – microphones and amplifiers allow us to detect even weaker sounds – as much as eight orders of magnitude weaker – and much stronger "sounds" called *shock waves*, produced by *supersonic* events such as explosions. The strongest supersonic "sounds" are some 32 orders of magnitude "louder" than the weakest sounds the human ear can detect, although even the weakest shock waves are almost strong enough to kill people.

It is very inconvenient to have to describe sound intensity in scientific notion across this wide a range – basically 40 orders of magnitude if not more (the limits of technology not being well defined at the low end of the scale). Also the human *mind* does not respond to sounds linearly. We do not psychologically perceive of sounds twice as intense as being twice as loud – in fact, a doubling of intensity is barely perceptible. Both of these motivate our using a different scale to represent sound intensities a relative *logarithmic* scale, called ***decibels***[267] .

The definition of a sound decibel is:

$$\beta = 10\log_{10}\left(\frac{I}{I_0}\right) \text{ dB} \tag{11.76}$$

where "dB" is the abbreviation for decibels (tenths of "bels", the same unit without the factor of 10). Note that $\log_{10}$ is log ***base 10***, not the natural log, in this expression. Also in this expression,

$$I_0 = 10^{-12} \text{ watts/meter}^2 \tag{11.77}$$

is the reference intensity, called the ***threshold of hearing***. It is, by definition, the "faintest sound the human ear can hear" although naturally Your Mileage May Vary here – the faintest sound *my* relatively old and deaf ears is very likely much louder than the faintest sound a young child can hear, and there obviously some normal variation (a few dB) from person to person at any given age.

The smallest increments of sound that the human ear can differentiate as being "louder" are typically ***two decibel increases*** – more likely 3. Let's see what a *doubling the intensity* does to the decibel level of the sound.

$$\begin{aligned}
\Delta\beta &= 10\left(\log_{10}\left(\frac{2I}{I_0}\right) - \log_{10}\left(\frac{I}{I_0}\right)\right) \\
&= 10\log_{10}\left(\frac{2I}{I_0}\frac{I_0}{I}\right) \\
&= 10\log_{10}(2) \\
&= 3.01dB
\end{aligned} \tag{11.78}$$

In other words, ***doubling the sound intensity from any value corresponds to an increase in sound intensity level of 3 dB!*** This is such a simple rule that it is religiously learned as a rule of thumb by engineers, physicists and others who have reason to need to work with intensities of any sort on a log (decibel) scale. 3 dB per factor of two in intensity can carry you a long, long way! Note well that we used one of the magic properties of logarithms/exponentials in this algebra (in case you are confused):

$$\log(A) - \log(B) = \log\left(\frac{A}{B}\right) \tag{11.79}$$

You should remember this; it will be very useful next year.

Table 7 presents a number of fairly common sounds, sounds you are likely to have directly or indirectly heard (if only from far away). Each sound is cross-referenced with the *approximate* peak ***overpressure*** P_0 in the sound pressure wave (in pascals), the sound intensity (in watts/meter^2), and the sound intensity level relative to the threshold of hearing in decibels.

[267] Wikipedia: http://www.wikipedia.org/wiki/Decibel. Note well that the term "decibel" is not restricted to sound – it is rather a way of transforming any quantity that varies over a very large (many orders of magnitude) range into a log scale. Other logarithmic scales you are likely to encounter include, for example, the Richter scale (for earthquakes) and the F-scale for tornadoes.

The overpressure is the pressure *over* the background of (a presumed) 1 atm in the sinusoidal pressure wave. The actual peak pressure would then be $P_{\max} = P_a + P_0$ while the minimum pressure would be $P_{\min} = P_a - P_0$. $P_{\min}$, however, ***cannot be negative*** as the lowest possible pressure is a vacuum, $P = 0$! At overpressures greater than 1 atm, then, it is no longer possible to have a pure sinusoidal sound wave. Waves in this category are a train of highly compressed peak amplitudes that drop off to (near) vacuum troughs in between, and are given their own special name: ***shock waves***. Shock waves are typically generated by very powerful phenomena and often travel faster than the speed of sound in a medium. Examples of shock waves are the sonic boom of a jet that has broken the sound barrier and the compression waves produced by sufficiently powerful explosives (close to the site of the explosion).

Shock waves as table 8 below clearly indicates, are capable of tearing the human body apart and accompany some of the most destructive phenomena in nature and human affairs – exploding volcanoes and colliding asteroids, conventional and nuclear bombs.

Source of Sound	P_0 (Pa)	dB
All sounds beyond this point are nonlinear ***shock waves***		
Human Death from Shock Wave Alone	200,000	200
1 Ton of TNT	632,000	210
Largest Conventional Bombs	2,000,000	220
1000 Tons of TNT	6,320,000	230
20 Kiloton Nuclear Bomb	63,200,000	250
57 Megaton (Largest) Nuclear Bomb	2,000,000,000	280
Krakatoa Volcanic Explosion (1883 C.E.)	63,200,000,000	310
Tambora Volcanic Explosion (1815 C.E.)	200,000,000,000	320

Table 8: Table of (approximate) P_0 and sound pressure levels in decibels relative to the threshold of human hearing at 10^{-12} watts/m^2 of ***shock waves***, events that produce distorted overpressures greater than one atmosphere. These "sounds" can be quite extreme! The Krakatoa explosion cracked a 1 foot thick concrete wall 300 miles away, was heard 3100 miles away, ejected 4 cubic miles of the earth, and created an audible pressure antinode on the opposite side of the earth. Tambora ejected 36 cubic miles of the earth and was equivalent to a *14 gigaton* nuclear explosion (14,000 1 megaton nuclear bombs)!

11.5: Doppler Shift

Everybody has heard the doppler shift in action. It is the rise (or fall) in frequency observed when a source/receiver pair approach (or recede) from one another. In this section we will derive expressions for the doppler shift for moving source and moving receiver.

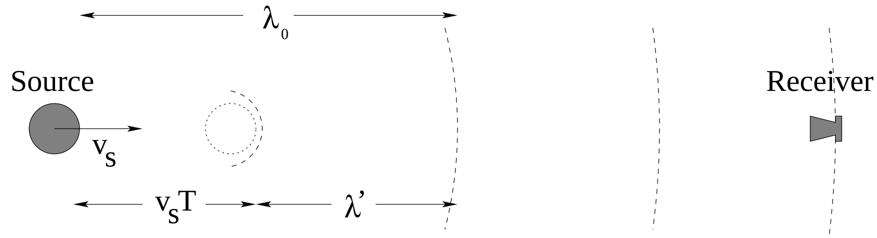


Figure 11.3: Waves from a source moving towards a stationary receiver have a foreshortened wavelength because the source moves ***in*** to the wave it produces. The *key* to getting the frequency shift is to recognize that the new (shifted) *wavelength* is $\lambda' = \lambda_0 - v_s T$ where T is the unshifted period of the source.

11.5.1: Moving Source

Suppose your receiver (ear) is stationary, while a source of harmonic sound waves at fixed frequency f_0 is approaching you. As the waves are emitted by the source they have a fixed wavelength $\lambda_0 = v_a/f_0 = v_a T$ and expand spherically from the point where the source was at the time the wavefront was emitted.

However, that point moves in the direction of the receiver. In the time between wavefronts (one period T) the source moves a distance $v_s T$. The shifted distance between successive wavefronts in the direction of motion (λ') can easily be determined from an examination of figure 11.3 above:

$$\lambda' = \lambda_0 - v_s T \tag{11.80}$$

We would really like the *frequency* of the doppler shifted sound. We can easily find this by using $\lambda' = v_a/f'$ and $\lambda = v_a/f$. We substitute and use $f = 1/T$:

$$\frac{v_a}{f'} = \frac{v_a - v_s}{f_0} \tag{11.81}$$

then we factor to get:

$$f' = \frac{f_0}{1 - \frac{v_s}{v_a}} \tag{11.82}$$

If the source is moving away from the receiver, everything is the same except now the wavelength is shifted to be bigger and the frequency smaller (as one would expect from changing the sign on the velocity):

$$f' = \frac{f_0}{1 + \frac{v_s}{v_a}} \tag{11.83}$$

11.5.2: Moving Receiver

Now imagine that the source of waves at frequency f_0 is stationary but the receiver is moving towards the source. The source is thus surrounded by spherical wavefronts a distance $\lambda_0 = v_a T$ apart. At $t = 0$ the receiver crosses one of them. At a time T' later, it has moved a distance $d = v_r T'$ in the direction of the source, and the wave from the source has moved a distance $D = v_a T'$ toward the receiver, and the receiver encounters the next wave front.

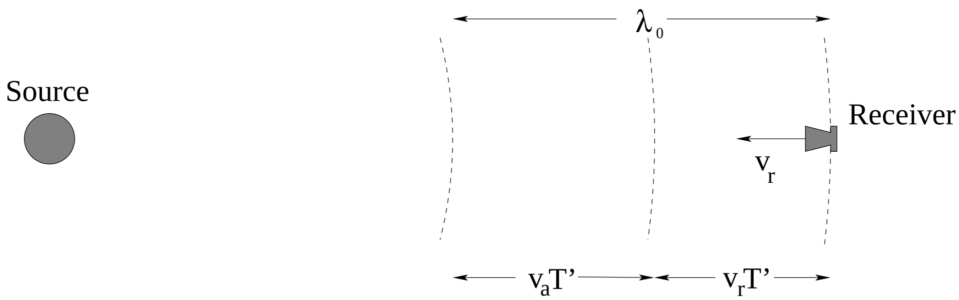


Figure 11.4: Waves from a stationary source are picked up by a moving receiver. They have a shortened ***period*** because the receiver doesn't wait for the next wavefront to reach it, at receives it when it has only moved part of a wavelength forward. The *key* to getting the frequency shift is to recognize that the sum of the distance travelled by the wave and the receiver in a new period T' must equal the original unshifted wavelength.

This can be visualized in figure 11.4 above. From it we can easily get:

$$\begin{aligned} \lambda_0 &= d + D && (11.84) \\ &= v_r T' + v_a T' && (11.85) \\ &= (v_r + v_a) T' && (11.86) \\ v_a T &= (v_r + v_a) T' && (11.87) \end{aligned}$$

We use $f_0 = 1/T$, $f' = 1/T'$ (where T' is the apparent time between wavefronts to the receiver) and rearrange this into:

$$f' = f_0 (1 + \frac{v_r}{v_a}) \tag{11.88}$$

Again, if the receiver is moving away from the source, everything is the same but the sign of v_r, so one gets:

$$f' = f_0 (1 - \frac{v_r}{v_a}) \tag{11.89}$$

11.5.3: Moving Source and Moving Receiver

This result is just the product of the two above – moving source causes one shift and moving receiver causes another to get:

$$f' = f_0 \frac{1 \mp \frac{v_r}{v_a}}{1 \pm \frac{v_s}{v_a}} \tag{11.90}$$

where in both cases *relative* approach shifts the frequency up and *relative* recession shifts the frequency down.

I do *not* recommend memorizing these equations – I don't have them memorized myself. It is *very* easy to confuse the forms for source and receiver, and the derivations take a few seconds and are likely worth points in and of themselves. If you're going to memorize anything, memorize the *derivation* (a process I call "learning", as opposed to "memorizing"). In fact, this is excellent advice for 90% of the material you learn in this course!

11.6: Standing Waves in Pipes

Everybody has created a stationary resonant harmonic sound wave by whistling or blowing over a beer bottle or by swinging a garden hose or by playing the organ. In this section we will see how to compute the harmonics of a given (simple) pipe geometry for an imaginary organ pipe that is open or closed at one or both ends.

The way we proceed is straightforward. Air cannot penetrate a closed pipe end. The air molecules at the very end are therefore “fixed” – they cannot displace into the closed end. The *closed* end of the pipe is thus a *displacement node*. In order *not* to displace air the closed pipe end has to exert a force on the molecules by means of pressure, so that the closed end is a pressure antinode.

At an open pipe end the argument is inverted. The pipe is open to the air (at fixed background/equilibrium pressure) so that there must be a pressure node at the open end. Pressure and displacement are $\pi/2$ out of phase, so that the *open* end is also a *displacement antinode*.

Actually, the air pressure in the standing wave doesn’t instantly equalize with the background pressure at an open end – it sort of “bulges” out of the pipe a bit. The displacement antinode is therefore just *outside* the pipe end, not *at* the pipe end. You may still draw a displacement antinode (or pressure node) as if they occur at the open pipe end; just remember that the distance from the open end to the first displacement node is *not* a very accurate measure of a quarter wavelength and that open organ pipes are a bit “longer” than they appear from the point of view of computing their resonant harmonics.

Once we understand the boundary conditions at the ends of the pipes, it is pretty easy to write down expressions for the standing waves and to deduce their harmonic frequencies.

11.6.1: Pipe Closed at Both Ends

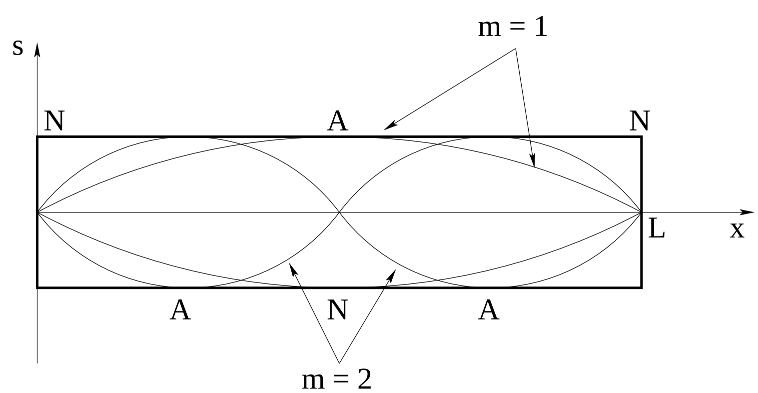


Figure 11.5: The pipe closed at both ends is ***just like a string fixed at both ends***, as long as one considers the ***displacement*** wave.

As noted above, we expect a ***displacement node*** (and hence ***pressure antinode*** at the closed end of a pipe, as air molecules cannot move through a solid surface. For a pipe closed at both ends, then, there are displacement nodes at both ends, as pictured above in figure 11.5.

This is just like a string fixed at both ends, and the solutions thus have the same functional form:

$$s(x,t) = s_0 \sin(k_m x)\cos(\omega_m t) \tag{11.91}$$

This has a node at $x = 0$ for all k. To get a node at the other end, we require (as we did for the string):

$$\sin(k_m L) = 0 \tag{11.92}$$

or

$$k_m L = m\pi \tag{11.93}$$

for $m = 1, 2, 3....$ This converts to:

$$\lambda_m = \frac{2L}{m} \tag{11.94}$$

and

$$f_m = \frac{v_a}{\lambda_m} = \frac{v_a m}{2L} \tag{11.95}$$

The $m = 1$ solution (first harmonic) is called the *principle harmonic* as it was before. The actual tone of a flute pipe with two closed ends will be a superposition of harmonics, usually dominated by the principle harmonic.

11.6.2: Pipe Closed at One End

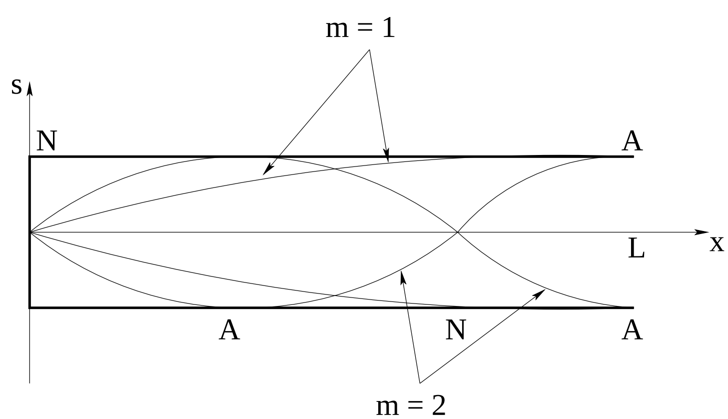


Figure 11.6: The pipe closed at both ends is ***just like a string fixed at one end***, as long as one considers the ***displacement*** wave.

In the case of a pipe open at only *one* end, there is a ***displacement node*** at the ***closed end***, and a ***displacement antinode*** at the ***open end***. If one considers the pressure wave, the positions of nodes and antinodes are ***reversed***. This is just like a string fixed at one end and free at the other. Let's arbitrarily make $x = 0$ the closed end. Then:

$$s(x,t) = s_0 \sin(k_m x)\cos(\omega_m t) \tag{11.96}$$

has a node at $x = 0$ for all k. To get an antinode at the other end, we require:

$$\sin(k_m L) = \pm 1 \tag{11.97}$$

or

$$k_m L = \frac{2m-1}{2}\pi \tag{11.98}$$

for $m = 1, 2, 3...$ (odd half-integral multiples of π. As before, you will see different conventions used to *name* the harmonics, with some books asserting that only odd harmonics are supported, but I prefer to make the harmonic index do exactly the same thing for both pipes so it counts the actual number of harmonics that *are* supported by the pipe. This is much more consistent with what one will do next semester considering e.g. interference, where one often encounters similar series for a phase angle in terms of odd-half integer multiples of π, and makes the second harmonic the lowest frequency actually present in the pipe in *all three cases* of pipes closed at neither, one or both ends.

This converts to:

$$\lambda_m = \frac{4L}{2m-1} \tag{11.99}$$

and

$$f_m = \frac{v_a}{\lambda_m} = \frac{v_a(2m-1)}{4L} \tag{11.100}$$

11.6.3: Pipe Open at Both Ends

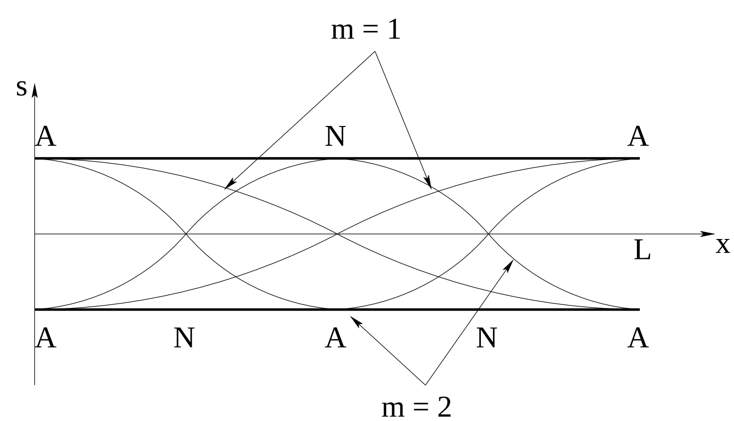


Figure 11.7: A pipe open at both ends is the exact *opposite* of a pipe (or string) closed (fixed) at both ends: It has *displacement antinodes* at the ends. Note well the principle harmonic with a single node in the center. The *resonant frequency* series for the pipe is the same, however, as for a pipe closed at both ends!

This is a ***panpipe***, one of the most primitive (and beautiful) of musical instruments. A panpipe is nothing more than a tube, such as a piece of hollow bamboo, open at both ends. The modes of this pipe are driven at resonance by blowing gently across one end, where the random fluctuations in the airstream are amplified only for the resonant harmonics.

To understand the frequencies of those harmonics, we note that there are ***displacement antinodes*** at ***both ends***. This is just like a string free at both ends. The displacement solution must thus be a *cosine* in order to have a displacement antinode at $x = 0$:

$$s(x,t) = s_0 \cos(k_m x)\cos(\omega_m t) \tag{11.101}$$

and

$$\cos(k_m L) = \pm 1 \qquad (11.102)$$

We can then write $k_m L$ as a series of suitable multiples of π and proceed as before to find the wavelengths and frequencies as a function of the mode index $m = 1, 2, 3....$ This is left as a (simple) exercise for you.

Alternatively, we could *also* note that there are *pressure* nodes at both ends, which makes them like a string fixed at both ends again as far as the *pressure* wave is concerned. This gives us *exactly* the same result (for frequencies and wavelengths) as the pipe closed at both ends above, although the pipe *open* at both ends is probably going to be a bit louder and easier to drive at resonance (how can you "blow" on a closed pipe to get the waves in there in the first place? How can the sound get out?

Either way one will get the same frequencies but the *picture* of the displacement waves is different from the *picture* of the pressure waves – be sure to draw displacement antinodes at the open ends if you are asked to draw a displacement wave, or vice versa for a pressure wave!

You might try drawing the first 2-3 harmonics on a suitable picture like the first two given above, with displacement antinodes at both ends. What does the principle harmonic look like? Show that the supported frequencies and wavelengths match those of the string fixed at both ends, or pipe closed at both ends.

11.7: Beats

If you have ever played around with a guitar, you've probably noticed that if two strings are fingered to be the "same note" but are really slightly out of tune and are struck together, the resulting sound "beats" – it modulates up and down in intensity at a low frequency often in the ballpark of a few cycles per second.

Beats occur because of the superposition principle. We can add any two (or more) solutions to the wave equation and still get a solution to the wave equation, even if the solutions have different frequencies. Recall the identity:

$$\sin(A) + \sin(B) = 2\sin(\frac{A+B}{2})\cos(\frac{A-B}{2}) \qquad (11.103)$$

If one adds two waves with different wave numbers/frequencies and uses this rule, one gets

$$\begin{aligned} s(x,t) &= s_0 \sin(k_0 x - \omega_0 t) + s_0 \sin(k_1 x - \omega_1 t) & (11.104)\\ &= 2s_0 \sin(\frac{k_0+k_1}{2}x - \frac{\omega_0+\omega_1}{2}t)\cos(\frac{k_0-k_1}{2}x - \frac{\omega_0-\omega_1}{2}t) & (11.105)\end{aligned}$$

This describes a wave that has twice the maximum amplitude, the *average* frequency (the first term), and a second term that (at any point x) oscillates like $\cos(\frac{\Delta\omega t}{2})$.

The "frequency" of this second modulating term is $\frac{f_0 - f_1}{2}$, but the ear *cannot hear* the inversion of phase that occurs when it is negative and the difference is small. It just hears maximum amplitude in the rapidly oscillating *average* frequency part, which goes to *zero* when the slowing varying cosine does, twice per cycle. The ear then *hears* two beats per cycle, making the "beat frequency":

$$f_{\rm beat} = \Delta f = |f_0 - f_1| \tag{11.106}$$

11.8: Interference and Sound Waves

We will not cover interference and diffraction of harmonic sound waves in this course. Beats are a common experience in sound as is the doppler shift, but sound wave interference is not so common an experience (although it can definitely and annoyingly occur if you hook up speakers in your stereo out of phase). Interference *will* be treated next semester in the context of coherent light waves. *Just* to give you a head start on that, we'll indicate the basic ideas underlying interference here.

Suppose you have two sources that are at the *same* frequency and have the *same* amplitude and phase but are at different locations. One source might be a distance x away from you and the other a distance $x + \Delta x$ away from you. The waves from these two sources add like:

$$\begin{aligned} s(x,t) &= s_0 \sin(kx - \omega t) + s_0 \sin(k(x + \Delta x) - \omega t) && (11.107)\\ &= 2s_0 \sin(k(x + \frac{\Delta x}{2} - \omega t) \cos(k\frac{\Delta x}{2}) && (11.108)\end{aligned}$$

The sine part describes a wave with twice the amplitude, the same frequency, but shifted slightly in phase by $k\Delta x/2$. The cosine part is *time independent* and *modulates* the first part. For some values of Δx it can *vanish*. For others it can have magnitude one.

The intensity of the wave is what our ears hear – they are insensitive to the phase (although certain echolocating species such as bats may be sensitive to phase information as well as frequency). The average intensity is proportional to the wave amplitude *squared*:

$$I_0 = \frac{1}{2}\rho\omega^2 s_0^2 v \tag{11.109}$$

With two sources (and a maximum amplitude of two) we get:

$$\begin{aligned} I &= \frac{1}{2}\rho\omega^2(2^2 s_0^2 \cos^2(k\frac{\Delta x}{2})v && (11.110)\\ &= 4I_0 \cos^2(k\frac{\Delta x}{2}) && (11.111)\end{aligned}$$

There are two cases of particular interest in this expression. When

$$\cos^2(k\frac{\Delta x}{2}) = 1 \tag{11.112}$$

one has *four times* the intensity of one source at peak. This occurs when:

$$k\frac{\Delta x}{2} = n\pi \tag{11.113}$$

(for $n = 0, 1, 2...$) or

$$\Delta x = n\lambda \tag{11.114}$$

If the *path difference* contains an *integral number of wavelengths* the waves from the two sources arrive *in phase*, add, and produce sound that has twice the amplitude and four times the intensity. This is called *complete constructive interference*.

On the other hand, when

$$\cos^2(k\frac{\Delta x}{2}) = 0 \tag{11.115}$$

the sound intensity *vanishes*. This is called *destructive* interference. This occurs when

$$k\frac{\Delta x}{2} = \frac{2n+1}{2}\pi \tag{11.116}$$

(for $n = 0, 1, 2...$) or

$$\Delta x = \frac{2n+1}{2}\lambda \tag{11.117}$$

If the path difference contains a *half integral* number of wavelengths, the waves from two sources arrive exactly out of phase, and *cancel*. The sound intensity vanishes.

You can see why this would make hooking your speakers up out of phase a bad idea. If you hook them up out of phase the waves *start* with a phase difference of π – one speaker is pushing out while the other is pulling in. If you sit equidistant from the two speakers and then harmonic waves with the same frequency from a single source coming from the two speakers *cancel* as they reach you (usually not perfectly) and the music sounds very odd indeed, because other parts of the music are not being played equally from the two speakers and don't cancel.

You can also see that there are many other situations where constructive or destructive interference can occur, both for sound waves and for other waves including water waves, light waves, even waves on strings. Our "standing wave solution" can be rederived as the superposition of a left- and right-travelling harmonic wave, for example. You can have interference from more than one source, it doesn't have to be just two.

This leads to some really excellent engineering. Ultrasonic probe arrays, radiotelescope arrays, sonar arrays, diffraction gratings, holograms, are all examples of interference being put to work. So it is worth it to learn the general idea as early as possible, even if it isn't assigned.

11.9: The Ear

Figure 11.8 shows a cross-section of the human ear, our basic transduction device for sound. This is not a biology course, so we will not dwell upon *all* of the structure visible in this picture, but rather will concentrate on the parts relevant to the physics.

Let's start with the ***outer ear***. This structure collects sound waves from a larger area than the ear canal per se and reflects them down to the ear canal. You can easily experiment with the kind of amplification that results from this by cupping your hands and holding them immediately behind your ears. You should be able to hear both a qualitative change in the frequencies you are hearing and an effective amplification of the sounds from in *front* of you at the expense of sounds originating *behind* you. Many animals have larger outer ears oriented primarily towards the front, and have muscles that permit them to further alter the direction

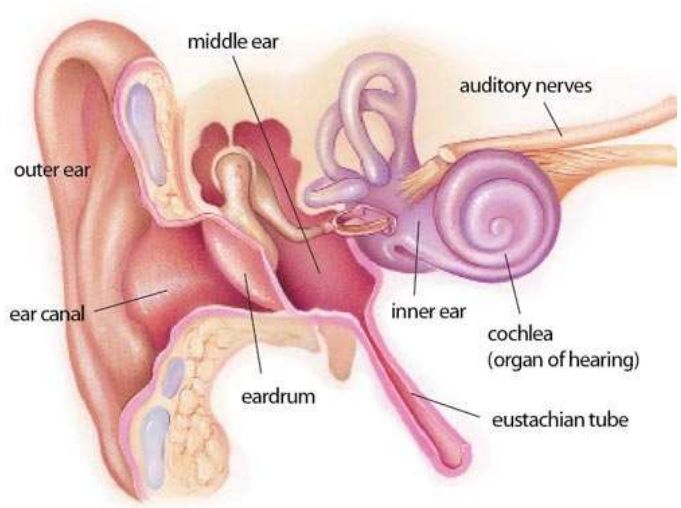


Figure 11.8: The anatomy of the human ear.

of most favorable sound collection without turning their heads. Human ears are more nearly omnidirection.

The ***auditory canal*** (ear canal in the figure above) acts like a resonant cavity to effectively amplify frequencies in the 2.5 kHz range and tune energy deliver to the tympanic membrane or eardrum. This membrane is a strong, resilient, tightly stretched structure that can vibrate in response to driving sound waves. It is connected to a collection of small bones (the ossicles) that conduct sound from the eardrum to the inner ear and that constitute the ***middle ear*** in the figure above. The common name of the ossicles are: hammer, anvil and stirrup, the latter so named because its shape strongly resembles that of the stirrup on a horse saddle. The anvil effectively amplifies oscillations by use of the principle of leverage, as a fulcrum attachment causes the stirrup end to vibrate through a much larger amplitude than the hammer end. The stirrup is directly connected to the ***oval window***, the gateway into the inner ear.

The middle ear is connected to the ***eustachian tube*** to your throat, permitting pressure inside your middle ear to equalize with ambient air pressure outside. If you pinch your nose, close your mouth, and try to breath out hard, you can actually blow air out through your ears although this is unpleasant and can be dangerous. This is one way your ears equilbrate by "popping" when you ride a car up a hillside or fly in an airplane. If/when this does not happen, pressure differences across the tympanic membrane reduce its response to ambient sounds reducing auditory acuity.

Sound, amplifed by focal concentration in the outer ear, resonance in the auditory canal, and mechanical leverage in the ossicles, enters the ***cochlea***, a shell-shaped spiral that is the primary organ of hearing that transduces sound energy into impulses in our nervous system through the oval window. The cochlea contains ***hair cells*** of smoothly varying length lining the narrowing spiral, each of which is ***resonant*** to a particular auditory frequency. The arrangement of the cells in a cross-section of the cochlea is shown in figure 11.9.

The nerves stretching from these cells are collected into the ***auditory nerve bundle*** and from thence carries the impulses they give off when they receive sound at the right frequency

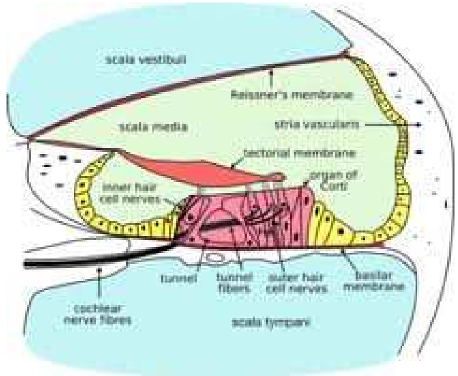


Figure 11.9: A cross-section of the spiral structure of the cochlea.

in to the ***auditory cortex*** (not shown) where it becomes, eventually and through a process still not fully understood, our perception of sound. Our brains take this frequency resolved information – the biomechanical equivalent of a fourier transform of the sound signal, in a way – and synthesize it back into a detailed perception of sound and music within the general frequency range of 10 Hz to 20,000 Hz.

As you can see, there are many individual parts that can fail in the human auditory system. Individuals can lose or suffer damage to their outer ears through accident or disease. The ear canal can become clogged with cerumen, or ***earwax***, a waxy fluid that normally cleans and lubricates the ear canal and eardrum but that can build up and dry out to both load the tympanic membrane so it becomes less responsive and physically occlude part of the canal so less sound energy can get through. The eardrum itself is vulnerable to sudden changes in sound pressure or physical contact that can puncture it. The middle ear, as a closed, warm, damp cavity connected to the throat, is an ideal breeding ground for certain bacteria that can cause infections and swelling that both interfere with or damage hearing and that can be quite painful. The ossicles are susceptible to physical trauma and infectious damage.

Finally, the hair cells of the cochlea itself, which are safely responsive over at least twelve to fourteen orders of magnitude of transient sound intensity (and safely responsive over eight or nine orders of magnitude of sustained sound intensity) are highly vulnerable to *both* sudden transient sounds of still higher intensity (e.g. sound levels in the vicinity of 120 to 140 decibels and higher *and* to sustained excitation at sound levels from roughly 90 decibels and higher. Both disease and medical conditions such as diabetes (that produces a progressive neuropathy) can further contribute to gradual or acute hearing loss at the neurological level.

When hair cells die, they do not regenerate and hearing loss of this sort is thus cumulative over a lifetime. It is therefore a really good idea to wear ear protectors if, for example, you play an instrument in a marching band or a rock and roll band where your hearing is routinely exposed to 100 dB and up sounds. It is also a good reason not to play music too loudly when you are young, however pleasurable it might seem. One is, after all, very probably trading listening to very loud music at age seventeen against listening to music *at all* at age seventy. Hearing aids do not really fix the problem, although they can help restore enough function for somebody to get by.

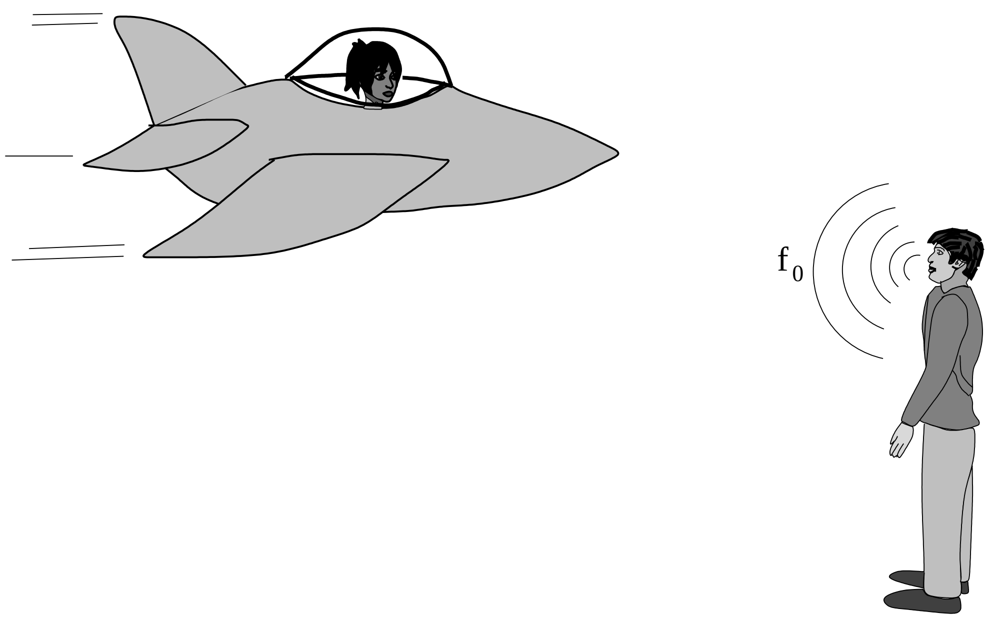
However, it is quite possible that over the next few decades the bright and motivated physics students of today will help create the bioelectronic and/or stem cell replacements of key organs and nervous tissue that will relegate age-related deafness to the past. I would certainly wish, as I sit here typing this with eyes and ears that are gradually failing as I age, that whether or not it come in time for me, it comes in time to help you.

Homework for Week 11

Problem 1.

Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

Problem 2.



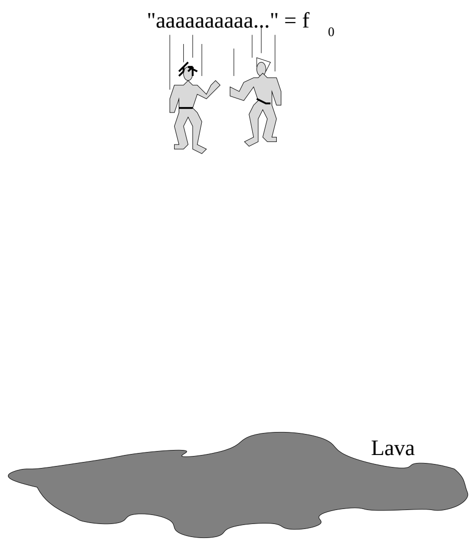
Fred is standing on the ground and Jane is blowing past him in a jet at a closest distance of approach of a few meters at twice the speed of sound in air. Both Fred and Jane are holding a loudspeaker that has been emitting sound at the *frequency* f_0 for some time.

a) Who hears the sound produced by the *other* person's speaker as *single frequency sound* when they are approaching one another and what frequency do they hear?

b) What does the *other* person hear (when they hear anything at all)?

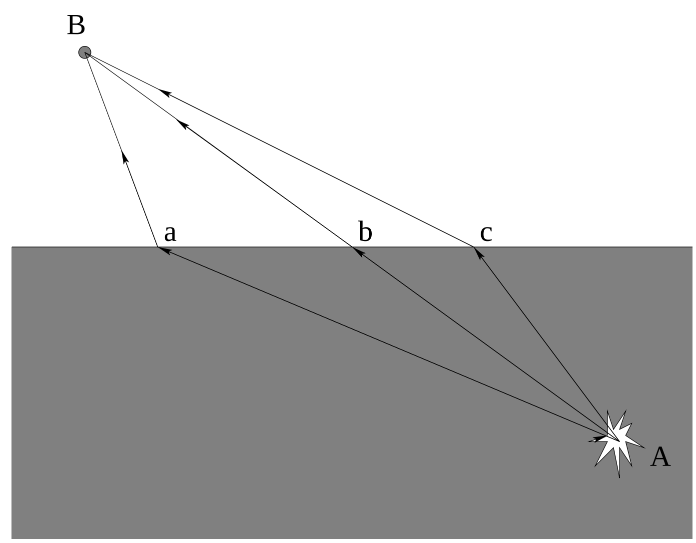
c) What frequenc(ies) do each of them hear *after* Jane has passed and is receding into the distance?

Problem 3.



Bill and Ted are falling into hell at a ***constant speed*** (terminal velocity), and are screaming at the frequency f_0. As they fall, they hear their own voices reflecting back to them from the puddle of molten rock that lies below at a frequency of $2f_0$. How fast are they falling relative to the speed of sound in warm, dry hellish air?

Problem 4.



Sound waves travel faster in water than they do in air. Light waves travel faster in air than they do in water. Based on this, which of the three paths pictured above are more likely to *minimize* the time required for the

a) Sound:

b) Light:

produced by an underwater explosion to travel from the explosion at A to the pickup at B? Why (explain your answers)?

Problem 5.

Discuss and answer the following questions using the concepts of *intensity* and *sound level in decibels*:

a) Sunlight reaches the surface of the earth with *roughly* 1000 Watts/meter^2 of intensity. What is the "sound intensity level" of a ***sound*** wave that carries as much ***energy per square meter***, in ***decibels***?

b) Using the table in the book, what kind of sound sources produce this sort of intensity? Bear in mind that the Sun is *150 million kilometers away* where sound sources capable of reaching the same intensity are typically only a few meters away. The the Sun produces a *lot* of (electromagnetic) energy compared to terrestrial sources of (sound) energy.

c) The human body produces energy at the rate of roughly 100 Watts. *Estimate* the fraction of this energy that goes into the actual sound of my lecture when I am speaking in a loud voice in front of the class (loud enough to be heard as loudly as normal conversation ten meters away).

d) Again using the table in the book, how far away from a jack hammer do you need to stand in order for the sound to (marginally) no longer be dangerous to your hearing?

e) You are listening to music and it isn't loud enough for you. You turn it up to ***quadruple*** the intensity! How much does the *sound level* of the music increase by? Does it matter what the sound level was before you turned it up?

Problem 6.

A long pipe is full of air at one atmosphere of pressure and at room temperature. A speaker at one end has inserted a travelling harmonic plane wave with wave number k and displacement amplitude s_0 into it:

$$s_1(x,t) = s_0 \sin(kx - \omega t)$$

A second long pipe is also full of air and has the superposition of two traveling harmonic plane waves moving in opposite directions inside:

$$s_2(x,t) = s_0 \sin(kx - \omega t) + 3s_0 \sin(kx + \omega t)$$

where k is the same as the first string.

If the average energy density in the first pipe $\left(\frac{dE_1}{dx}\right)_{\rm avg}$, what is the average energy density of the second string $\left(\frac{dE_2}{dx}\right)_{\rm avg}$ in *terms* of $\left(\frac{dE_1}{dx}\right)_{\rm avg}$.

Problem 7.

Two identical strings of length L have mass μ and are fixed at both ends. One string has tension T. The other has tension $1.21T$. When plucked, the first string produces a tone at frequency f_1, the second produces a tone at frequency f_2.

a) What is the *beat* frequency produced if the two strings are is plucked at the same time, in terms of f_1?

b) Are the beats likely to be audible if f_1 is 500 Hz? How about 50 Hz? Why or why not?

Problem 8.

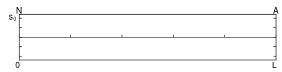
The following short answer questions are intended to help you solidify the ***scaling*** of sound intensity/sound level as various parameters are varied. Practice them until you are skilled and fully understand what goes into them.

a) You measure the sound intensity level of a single frequency sound wave produced by a loudspeaker with a calibrated microphone to be 80 dB. At that intensity, the *peak overpressure* (amplitude of the pressure oscillation *around* P_a, the background atmospheric pressure) in the sound wave at the microphone is P_0. The loudspeaker's amplitude is turned up until the intensity level is 100 dB. What is the peak overpressure P_1 of the sound wave now ***in terms of*** P_0?

b) You see some friends of yours on the other side of a ravine 1 kilometer across. You attempt to halloo them to signal them to meet you down by the lake later. The maximum intensity of your yelling voice is 10^{-2} w/m^2, measured one meter away from your mouth. Is it likely that your friends will be able to hear (barely) you? Assume that their local environment is relatively quiet at around 40 dB.

c) A certain speaker emits sound with frequency ω_0 and peak displacement s_0 at an intensity of 0.1 W/m^2 right in front of the speaker. If the frequency of the speaker is doubled at *constant power*, what happens to the peak displacement?

Problem 9.



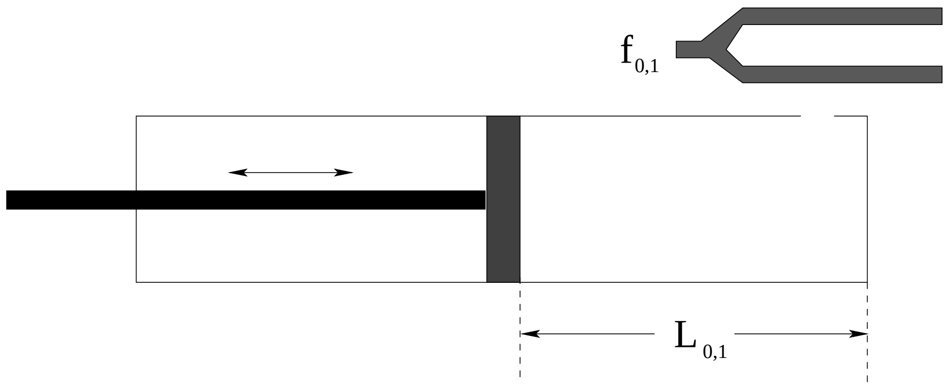
An organ pipe is made from a brass tube closed at one end as shown. The pipe has length L and the speed of sound in air is v_a. When played, it produces a sound that is a mixture of the ***first, fifth and seventh harmonics***:

a) What are the frequencies of these harmonics?

b) Qualitatively sketch the ***displacement wave amplitude envelopes*** for these three harmonics in on (copies of) the provided axes, indicating the nodes and antinodes.

c) Evaluate your answers for the frequencies numerically when $L = 3.4$ meters long, and $v_a = 340$ meters/second (as usual).

Problem 10.



You crash land on a strange planet and all your apparatus for determining if the planet's atmosphere is like Earth's is wrecked. In desperation you decide to measure the speed of sound in the atmosphere before taking off your helmet. You do have a barometer handy and can see that the air pressure outside is approximately one atmosphere and the temperature seems to be about 300 °K, so if the speed of sound is the same as on Earth the air *might* be breathable.

You jury rig a piston and cylinder arrangement like the one shown above (where the cylinder is ***closed at both ends*** but has a small hole in the side to let sound energy in to resonate) and take out your two handy tuning forks, one at $f_0 = 3400$ Hz and one at $f_1 = 6800$ Hz.

a) Using the 3400 Hz fork as shown, what do you *expect* (or rather, hope) to hear as you move the piston in and out (varying L_0). In particular, what are the shortest few values of L_0 for which you expect to hear a maximum resonant intensity from the tube if the speed of sound in the unknown atmosphere is indeed the same as in air (which you will cleverly note I'm *not* telling you as you are supposed to know this number)?

b) Using the 6800 Hz fork you hear your first maximum (for the smallest value of L_1) at $L_1 = 5$ cm. Should you sigh with relief and rip off your helmet?

c) What is the *next* value for L_1 for which you should hear a maximum (given the measurement in b) and what should the difference between the two equal in terms of the wavelength of the 6800 Hz wave in the unknown gas? Draw the *displacement* wave for this case only schematically in on the diagram above (assuming that the L shown is this second-smallest value of L_1 for the f_1 tuning fork), and indicate where the nodes and antinodes are.

Final Exam AND Second Hour Exam!

Study for the second hour exam and the final exam! We are starting the last week of class next week, and this wraps up both the chapter and the textbook! Students looking for *more* problems to work on are directed to the online review guide for introductory physics 1 and the online math review as time permits or as needed. But ***first*** I ***strongly recommend*** that you work on ***mastering the many, many problems you've had where you now have access to the (often richly annotated!) solutions!***

Additional generic advice: When going over problems you've solved (or when looking at new ones) ***practice starts!*** That is, look at the problem, identify the physics needed, and perhaps write down the *starting* equations or concepts. Then tell yourself *what you would do to solve it the rest of the way* (without actually doing it). As you master the concepts of the problem(s), you'll get to where you see a problem, recognize the physics, get the physics correctly *started* on the page and then – you just follow your nose to solve it, doing the algebra and rearrangements as needed. The ***really hard part is getting it started!*** Algebra, if that's all that is left, is comparatively easy and heck, I might even *help* you if you get stuck on the *algebra* on an exam, but I cannot help you with the physics! Of course, periodically you should pick a problem and solve it all the way just for the additional practice.

Practicing starts lets you cover 2-3 times as many problems in any given block of study time, and feel better about the result at the end of it as well. It can give you the *confidence* to know that at least, you won't fail or get an embarrassingly low score on the exam(s), and it has an excellent chance of producing an embarrassingly *high* score, a *gaudy* score if you are careful with the algebra and problem *finishes*.

Bear in mind that our grading policy is to usually give at least half credit for having the right physics down in the right way, even if the rest of the math is a mess. And if you do get it down on the paper, and don't panic, and do remember solving the problems *like* that problem on quizzes, exams, homework, in-class problem sets, and worked examples in lecture and the textbook, you certainly have a good chance of finishing it right!

Week 12: Gravity

1.19: Gravity Summary

- Early western (Greek) cosmology was both ***geocentric*** – simple earth-centered model with fixed stars "lamps" or "windows" on big solid bowl, moon and stars and planets orbiting the (usually flat) Earth "somehow" in between. The simple geocentric models failed to explain ***retrograde motion*** of the planets, where for a time they seem to go backwards against the fixed stars in their general orbits. There were also early *heliocentric* – sun centered – models, in particular one by Aristarchus of Samos (270 B.C.E.), who used ***parallax*** to measure the size of the earth and the sizes of and distances to the Sun and Moon.

- Ptolemy[268] (140 C.E.) "explained" retrograde motion with a ***geometric geocentric model*** involving complex ***epicycles***. Kudos to Ptolemy for inventing geometric modelling in physics! The model was a genuine scientific hypothesis, in principle ***falsifiable***, and a good starting place for further research.

  Sadly, a few hundred years later the state religion of the western world's largest empire embraced this geocentric model as being ***consistent with The Book of Genesis*** in its theistic scriptural mythology (and with many other passages in the old and new testaments) and for over a thousand years alternative explanations were considered heretical and could only be made at substantial personal risk throughout the Holy Roman Empire.

- Copernicus[269] (1543 C.E.) (re)invented a ***heliocentric*** – sun-centered model, explained retrograde motion with ***simpler*** plain circular geometry, regular orbits. The work of Copernicus, *De Revolutionibus Orbium Coelestium*[270] (On the Revolutions of the Heavenly Spheres) was forthwith banned by the Catholic Church as heretical at the same time that Galileo was both persecuted and prosecuted.

- Wealthy Tycho Brahe accumulated data and his paid assistant, Johannes Kepler, fit that data to specific orbits and deduced ***Kepler's Laws***. All Brahe got for his efforts was a lousy moon crater named after *him*[271] .

- **Kepler's Laws:**

  a) All planets move in elliptical orbits with the sun at one focus.

[268] Wikipedia: http://www.wikipedia.org/wiki/Ptolemy.
[269] Wikipedia: http://www.wikipedia.org/wiki/Copernicus.
[270] Wikipedia: http://www.wikipedia.org/wiki/De_revolutionibus_orbium_coelestium.
[271] Wikipedia: http://www.wikipedia.org/wiki/Tycho_(crater).

b) A line joining any planet to the sun sweeps out equal areas in equal times (dA/dt = constant).

c) The square of the period of any planet is proportional to the cube of the planet's mean distance from the sun ($T^2 = CR^3$). Note that the semimajor or semiminor axis of the ellipse will serve as well as the mean, with different contants of proportionality.

- Galileo[272] (1564-1642 C.E.) is known as the Copernican heliocentric model's most famous early defender, not so much because of the quality of his science as for his infamous ***prosecution by the Catholic church***. In truth, Galileo was a contemporary of Kepler and his work was nowhere nearly as carefully done or mathematically convincing (or correct!) as Kepler's, although using a ***telescope*** he made a number of important discoveries that added considerable further weight to the argument in favor of heliocentrism in general.

- Newton[273] (1642-1727 C.E.) was the inheritor of the tremendous advances of Brahe, Descartes[274] (1596-1650 C.E.), Kepler, and Galileo. Applying the analytic geometry invented by Descartes to the empirical laws discovered by Kepler and the kinematics invented by Galileo, he was able to deduce ***Newton's Law of Gravitation***:

$$\vec{\boldsymbol{F}} = -\frac{GMm}{r^2}\hat{\boldsymbol{r}} \tag{12.1}$$

(a simplified form valid when mass $M \gg m$, $\vec{\boldsymbol{r}}$ are coordinates centered on the larger mass M, and $\vec{\boldsymbol{F}}$ is the force acting on the smaller mass); we will learn a more precisely stated version of this law below. This law fully explained at the limit of observational resoution, and continues to ***mostly*** explain, Kepler's Laws and the motions of the planets, moons, comets, and other visible astronomical objects! Indeed, it allows their orbits to be precisely computed and extrapolated into the distant past or future from a sufficient knowledge of initial state.

- In Newton's Law of Gravitation the constant G is a considered to be a ***constant of nature***, and was measured by Cavendish[275] in a famous experiment, thus (as we shall see) "weighing the planets". The value of G we will use in this class is:

$$G = 6.67 \times 10^{-11} \quad \frac{\text{N-m}^2}{\text{kg}^2} \tag{12.2}$$

You are responsible for knowing this number! Like g, it is enormously important and useful as a key to the relative strength of the forces of nature and explanation for why it takes an entire planet to produce a force on your body that is easily opposed by (for example) a thin nylon rope.

- The ***gravitational field*** is a simplification of Newton's theory of gravitation that emerged over a considerable period of time with no clear author that attempts to resolve the problem Newton first addressed of ***action at a distance*** – the need for a ***cause*** for the gravitational force that propagates ***from*** one object ***to*** the other. Otherwise it is difficult

[272] Wikipedia: http://www.wikipedia.org/wiki/Galileo.
[273] Wikipedia: http://www.wikipedia.org/wiki/Newton.
[274] Wikipedia: http://www.wikipedia.org/wiki/Descartes.
[275] Wikipedia: http://www.wikipedia.org/wiki/Cavendish.

to understand how one mass "knows" of the mass, direction and distance of its partner in the gravitational force! It is (currently) defined to be the ***gravitational force per unit mass*** or ***gravitational acceleration*** produced at and associated with every ***point in space*** by a ***single*** massive object. This field acts on any mass placed at that point and thereby exerts a force. Thus:

$$\vec{\boldsymbol{g}}(\vec{\boldsymbol{r}}) = -\frac{GM}{r^2}\hat{\boldsymbol{r}} \tag{12.3}$$

$$\vec{\boldsymbol{F}}_m(\vec{\boldsymbol{r}}) = m\vec{\boldsymbol{g}}(\vec{\boldsymbol{r}}) = -\frac{GMm}{r^2}\hat{\boldsymbol{r}} \tag{12.4}$$

- Important true facts about the gravitational field:
  - The gravitational field produced by a (thin) spherically symmetric shell of mass ΔM vanishes inside the shell.
  - The gravitational field produced by this same shell equals the usual

    $$\vec{\boldsymbol{g}}(\vec{\boldsymbol{r}}) = -\frac{G\Delta M}{r^2}\hat{\boldsymbol{r}} \tag{12.5}$$

    outside of the shell. As a consequence the field outside of any spherically symmetric distribution of mass is just

    $$\vec{\boldsymbol{g}}(\vec{\boldsymbol{r}}) = -\frac{G\Delta M}{r^2}\hat{\boldsymbol{r}} \tag{12.6}$$

  These two results can be proven by direct integration or by using Gauss's Law for the gravitational field (using methodology developed next semester for the electrostatic field).

- The gravitational force is conservative. The gravitational potential energy of mass m in the field of mass M is:

  $$U_m(\vec{\boldsymbol{r}}) = -\int_{\infty}^{\vec{\boldsymbol{r}}} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{\ell}} = -\frac{GMm}{r} \tag{12.7}$$

  By convention, the zero of gravitational potential energy is at $r_0 = \infty$ (in all directions).

- The ***gravitational potential*** is to the potenial energy as the gravitational field is to the force. That is:

  $$V(\vec{\boldsymbol{r}}) = \frac{U_m(\vec{\boldsymbol{r}})}{m} = -\int_{\infty}^{\vec{\boldsymbol{r}}} \vec{\boldsymbol{g}} \cdot d\vec{\boldsymbol{\ell}} = -\frac{GM}{r} \tag{12.8}$$

  It as a scalar field that depends only on distance, it is the simplest of the ways to describe gravitation. Once the potential is known, one can always find the gravitational potential energy:

  $$U_m(\vec{\boldsymbol{r}}) = mV(\vec{\boldsymbol{r}}) \tag{12.9}$$

  or the gravitational field:

  $$\vec{\boldsymbol{g}}(\vec{\boldsymbol{r}}) = -\vec{\boldsymbol{\nabla}}V(\vec{\boldsymbol{r}}) \tag{12.10}$$

  or the gravitational force:

  $$\vec{\boldsymbol{F}}_m(\vec{\boldsymbol{r}}) = -m\vec{\boldsymbol{\nabla}}V(\vec{\boldsymbol{r}}) = m\vec{\boldsymbol{g}}(\vec{\boldsymbol{r}}) \tag{12.11}$$

- ***Escape velocity*** is the minimum velocity required to escape from the surface of a planet (or other astronomical body) and coast in free-fall all the way to infinity so that the object "arrives at infinity at rest". Since $U(\infty) = 0$ by definition, the ***escape energy*** for a particle is:
$$E_{\rm escape} = K(\infty) + U(\infty) = 0 + 0 = 0 \tag{12.12}$$
Since mechanical energy is conserved moving through the (presumed) vacuum of space, the total energy must be zero on the surface of the planet as well, or:
$$\frac{1}{2}mv_e^2 - \frac{GMm}{R} = 0 \tag{12.13}$$
or
$$v_e = \sqrt{\frac{2GM}{R}} \tag{12.14}$$
On the earth:
$$v_e = \sqrt{\frac{2GM}{R}} = \sqrt{2gR_e} = 11.2 \times 10^3 \text{ meters/second} \tag{12.15}$$
(11.2 kilometers per second). This is also the most reasonable starting estimate for the speed with which falling astronomical objects, e.g. meteors or asteroids, will ***strike*** the earth. A large falling mass loses basically all of its kinetic energy on impact, so that even a fairly small asteroid can easily strike with an explosive power greater than that of a ***nuclear bomb***, or ***many nuclear bombs***. It is believed that just such a collision was responsible for at least the final Cretaceous extinction event that brought an end to the age of the dinosaurs some sixty million years ago, and similar collisions may have caused other great extinctions as well.

- A (point-like) object in a plane orbit has a kinetic energy that can be written as:
$$K = K_{\rm rot} + K_r = \frac{L^2}{2mr^2} + \frac{1}{2}mv_r^2 \tag{12.16}$$
The total mechanical energy of this object is thus:
$$E = K + U = \frac{1}{2}mv_r^2 + \frac{L^2}{2mr^2} - \frac{GMm}{r} \tag{12.17}$$
$\vec{L}$ for an orbit (in a central force, recall) is ***constant***, hence L^2 is constant in this expression. The total energy and the angular momentum thus become convenient ways to parameterize the orbit.

- The ***effective potential energy*** is of a mass m in an orbit with (magnitude of) angular momentum L is:
$$U'(r) = \frac{L^2}{2mr^2} - \frac{GMm}{r} \tag{12.18}$$
and the total energy can be written in terms of the ***radial kinetic energy only*** as:
$$E = \frac{1}{2}mv_r^2 + U'(r) \tag{12.19}$$
This is a convenient form to use to make ***energy diagrams*** and determine the ***radial turning points of an orbit***, and permits us to easily classify orbits not only as ellipses but as general ***conic sections***. The term $L^2/2mr^2$ is called the ***angular momentum barrier*** because it's negative derivative with respect to r can be interpreted as a strongly (radially) repulsive pseudoforce for small r.

- The orbit classifications (for a given nonzero L) are:
  - Circular: Minimum energy, only one permitted value of r_c in the energy diagram where $E = U'(r_c)$.
  - Elliptical: Negative energy, always have two turning points.
  - Parabolic: Marginally unbound, $E = 0$, one radial turning point. This is the "escape orbit" described above.
  - Hyperbolic: Unbound, $E > 0$, one radial turning point. This orbit has enough energy to reach infinity while still moving, if you like, although a better way to think of it is that its asymptotic radial kinetic energy is greater than zero.

12.1: Cosmological Models

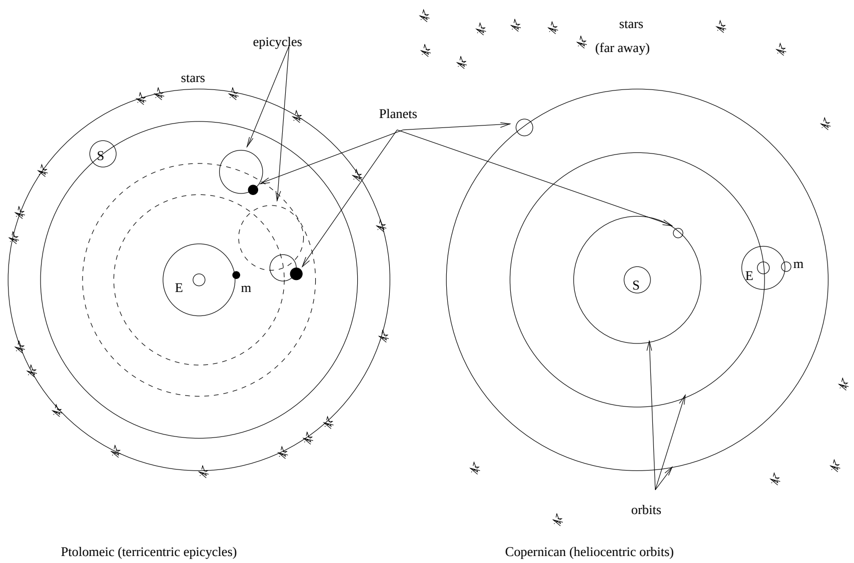


Figure 12.1: The Ptolemaic ***geocentric*** model with ***epicycles*** that sufficed to explain the observational data of retrograde motion. The Copernican ***geocentric*** model also explained the data and was somewhat simpler. To determine which was correct required the use of ***parallax*** to determine *distances* as well as angles.

Early western (Greek) cosmology was both ***geocentric***, with fixed stars "lamps" or "windows" on a big solid bowl, the moon and sun and planets orbiting a fixed, stationary Earth in the center. Plato represented the Earth (approximately correctly) as a sphere and located it at the center of the Universe. Astronomical objects were located on transparent "spheres" (or circles) that rotated uniformly around the Earth at differential rates. Euxodus and then Aristotle (both students of Plato) elaborated on Plato's original highly idealized description, adding spheres until the model "worked" to some extent, but left a number of phenomena either unexplained or (in the case of e.g. lunar phases) not particularly *believably* explained.

The principle failure of the Aristotelian geocentric model is that it fails to explain ***retrograde motion*** of the planets, where for a time they seem to go backwards against the fixed stars in their general orbits. However, in the second century Claudius Ptolemaeus constructed a somewhat simpler geocentric model that is currently known as the ***Ptolemaic model*** that still involved Plato's circular orbits with stars embedded on an outer revolving sphere, but added to this the notion of *epicycles* – planets orbiting in circles *around* a point that was itself in a circular orbit around the Earth. The model was very complex, but it actually *explained the observational data including retrograde motion* well enough that – for a variety of political, psychological, and religious reasons – it was adopted as ***the "official" cosmology of Western Civilization, endorsed and turned into canonical dogma by the Catholic Church*** as geocentrism agreed (more or less) with the cosmological assertions of the Bible.

In this original period – during which the Greeks invented things like mathematics and philosophy and the earliest rudiments of physics – geocentrism was not the only model. The Pythogoreans, for example, postulated that the earth orbited a "great circle of fire" that was always beneath one's feet in a flat-earth model, while the sun, stars, moon and so on orbited the whole thing. An "anti-Earth" was supposed to orbit on the far side of the great fire, where we cannot see it. All one can say is gee, they must have had really good recreational/religious hallucinogenic drugs back then...[276]. Another "out there" model – by the standards of the day – was the ***heliocentric*** model.

The first person known to have proposed a heliocentric system, however, was Aristarchus of Samos (c. 270 BC). Like Eratosthenes, Aristarchus calculated the size of the Earth, and measured the size and distance of the Moon and Sun, in a treatise which has survived. From his estimates, he concluded that the Sun was six to seven times wider than the Earth and thus hundreds of times more voluminous. His writings on the heliocentric system are lost, but some information is known from surviving descriptions and critical commentary by his contemporaries, such as Archimedes. Some have suggested that his calculation of the relative size of the Earth and Sun led Aristarchus to conclude that it made more sense for the Earth to be moving than for the huge Sun to be moving around it.

Archimedes was familiar with, and apparently endorsed, this model. This model explained the lack of motion of the stars by putting them ***very far away*** so that distances to them could not easily be detected using parallax! This was the first hint that the Universe was ***much larger than geocentric models assumed***, which eliminated any *need* for parallax by approximately fixing the earth itself relative to the stars.

The heliocentric model explained may things, but it wasn't clear how it would explain (in particular) retrograde motion. For a variety of reason (mostly political and religious) the platonic geocentric model was preserved and the heliocentric model officially forgotten and ignored until the early 1500's, when a catholic priest and polymath[277] resurrected it and showed how it *explained* retrograde motion with far less complexity than the Ptolemaic model. Since the work of Aristarchus was long forgotten, this reborn heliocentric model was called the ***Copernican model***[278] , and was perhaps the spark that lit the early Enlightenment[279] .

Initially, the Copernican model, published in 1543 by a Copernicus who was literally on his deathbed, attracted little attention. Over the next 70 years, however, it gradually caused more and more debate, in no little part because it directly contradicted a number of passages in the Christian holy scriptures and thereby strengthened the position of an increasing number of contemporary philosophers who challenged the divine inspiration and fidelity of those writings. This drew attention from scholars within the established Catholic church as well as from the new Protestant churches that were starting to emerge, as well as from other philosophers.

The most important of these philosophers was another polymath by the name of Galileo

---

[276] Which in fact, they did...

[277] One who is skilled at many philosophical disciplines. Copernicus made contributions to astronomy, mathematics, medicine, economics, spoke four languages, and had a doctorate in law.

[278] Wikipedia: http://www.wikipedia.org/wiki/Copernican heliocentrism.

[279] Wikipedia: http://www.wikipedia.org/wiki/Age of Enlightenment. The Enlightenment was the philosophical revolution that led to the invention of physics and calculus as the core of "natural philosophy" – what we now call science – as well as economics, democracy, the concept of "human rights" (including racial and sexual equality) within a variety of social models, and to the rejection of scriptural theism as a means to knowledge that had its roots in the discoveries of Columbus (that the world was not flat), Descartes (who invented analytic geometry), Copernicus (who proposed that the non-flat Earth was not the center of creation after all), setting the stage in the sixteenth century for radical and rapid change in the seventeenth and eighteenth centuries.

Galilei[280] . The first refracting telescopes were built by spectacle makers in the Netherlands in 1608; Galileo heard of the invention in 1609 and immediately built one of his own that had a magnification of around 20. With this instrument (and successors also of his own design) he performed an amazing series of astronomical observations that permitted him to *empirically support* the Copernican model in preference over the Ptolemaic model.

It is important to note well that *both* models explain the observations available to the naked eye. Ptolemaeus' model was somewhat more complex than the Copernican model (which weighs against it) but one common early complaint against the latter was that it wasn't provable by observation and all of the sages and holy fathers of the church for nealy 2000 years considered geocentrism to be true on observational grounds.

Galileo's telescope – which was little more powerful than an ordinary pair of hand-held binoculars today – was sufficient to provide that proof. Galileo's instrument clearly revealed that the moon was a *planetoid object*, a truly massive ball of rock that orbited the Earth, so large that it had its own mountains and "seas". It revealed that Jupiter had not one, but four similar moons of its own that orbited *it* in similar manner (moons named "The Galilean Moons" in his honor). He observed the phases of Venus as it orbited the sun, and correctly interpreted this as positive evidence that Venus, too, was a huge world orbiting the sun as the Earth orbits the sun while revolving and being orbited by its own moon. He was one of the first individuals in modern times to observe sunspots (although he incorrectly interpreted them as Mercury in transit across the Sun) and set the stage for centuries of solar astronomical observations and sunspot counts that date from roughly this time. His (independent) observations on gravity even helped inspire Newton to develop gravity as the *universal cause* of the observed orbital motions.

However, the publication of his own observations defending Copernicus corresponded almost exactly with the Church finally taking action to condemn the work of Copernicus and ban his book describing the model. In 1600 the Roman Inquisition had found Catholic priest, freethinker, and philosopher Giordano Bruno[281] guilty of heresy and burned him at the stake, establishing a dangerous precedent that put a damper on the development of science everywhere that the Roman church held sway.

Bruno not only embraced the Copernican theory, he went far beyond it, recognizing that the Sun is a star like other stars, that there were far, far more stars than the human eye could see without help, and he even asserted that many of those stars have planets like the Earth and that those planets were likely to be inhabited by intelligent beings. While Galileo was aided in his assertions by the use of the telescope, Bruno's were all the more remarkable because they *preceded* the invention of the telescope. Note well that the human eye can only make out some ***3500 stars altogether*** unaided on the darkest, clearest nights. This leap from 3500 to "infinity", and the other inferences he made to accompany them, were quite extraordinary. His guess that the stars are effectively numberless was validated shortly afterwards by means of the very first telescopes, which revealed more and more stars in the gaps between the visible stars as the power of the telescopes was systematically increased.

[280] Wikipedia: http://www.wikipedia.org/wiki/Galileo Galilei. It would take too long to recite all of Galileo's discoveries and theories, but Galileo has for good reason been called "The Father of Modern Science".

[281] Wikipedia: http://www.wikipedia.org/wiki/Giordano Bruno. Bruno is, sadly, almost unknown as a philosopher and early scientist for all that he was braver and more honest in his martyrdom that Galileo in his capitulation.

We only discovered positive evidence of the first confirmed exoplanet[282] in *1988* and are *still* in the process of searching for evidence that might yet validate his further hypothesis of life spread throughout the Universe, some of it (other than our own) intelligent. Galileo had written a letter to Kepler in 1597, a mere three years before Bruno's ritualized murder, stating his belief in the Copernican system (which was not, however, the direct cause of Bruno's conviction for heresy). The stakes were indeed high, and piled higher still with wood.

Against this background, Galileo developed a careful and observationally supported argument in favor of the Copernican model and began cautiously to publish it within the limited circles of philosophical discourse available at the time, proposing it as a "theory" only, but arguing that it did not contradict the Bible. This finally attracted the attention of the church. Cardinal and Saint Robert Bellarmine wrote a famous letter to Galileo in 1615[283] explaining the Church's position on the matter. This letter should be required reading for all students, and since if you are reading this textbook you are, in a manner of speaking, *my* student, please indulge me by taking a moment and following the link to read the letter and some of the commentary following.

In it Bellarmine makes the following points:

- If Copernicus (and Galileo, defending Copernicus and advancing the theory in his own right) are correct, the heliocentric model "is a very dangerous thing, not only by irritating all the philosophers and scholastic theologians, but also by injuring our holy faith and rendering the Holy Scriptures false."

  In other words, if Galileo is correct, the holy scriptures are incorrect. Bellarmine correctly infers that this would reduce the degree of belief in the infallibility of the holy scriptures and hence the entire basis of belief in the religion they describe.

- Furthermore, Bellarmine continues, Galileo is *disagreeing with established authorities* with his hypothesis, who "...all agree in explaining literally (ad litteram) that the sun is in the heavens and moves swiftly around the earth, and that the earth is far from the heavens and stands immobile in the center of the universe. Now consider whether in all prudence the Church could encourage giving to Scripture a sense contrary to the holy Fathers and all the Latin and Greek commentators. Nor may it be answered that this is not a matter of faith, for if it is not a matter of faith from the point of view of the subject matter, it is on the part of the ones who have spoken."

- Finally, Bellarmine concludes that "if there were a true demonstration that the sun was in the center of the universe and the earth in the third sphere, and that the sun did not travel around the earth but the earth circled the sun, then it would be necessary to proceed with great caution in explaining the passages of Scripture which seemed contrary, and we would rather have to say that we did not understand them than to say that something was false which has been demonstrated." He goes on to assert that "the words 'the sun also riseth and the sun goeth down, and hasteneth to the place where he ariseth, etc.' were those of Solomon, who not only spoke by divine inspiration but was a man

[282] Wikipedia: http://www.wikipedia.org/wiki/Extrasolar planet. As of today, some 851 planets in 670 systems have been discovered, with more being discovered almost every day using a dazzling array of sophisticated techniques.
[283] http://www.fordham.edu/halsall/mod/1615bellarmine-letter.asp

wise above all others and most learned in human sciences and in the knowledge of all created things, and his wisdom was from God."

Interested students are invited to play *Logical Fallacy Bingo*[284] with the text of the entire document. Opinion as fact, appeal to consequences, wishful thinking, appeal to tradition, historian's fallacy, argumentum ad populum, thought-terminating cliche, and more. The argument of Bellarmine boils down to the following:

- If the heliocentric model is true, the Bible is false where that model contradicts it.
- If the Bible is false *anywhere*, it cannot be trusted *everywhere* and Christianity itself can legitimately be doubted.
- The Bible and Christianity are true. Even if they appear to be false they are *still* true, but don't worry, they don't even appear to be false.
- Therefore, while it is all very well to show how a heliocentric model could *mathematically, or hypothetically* explain the observational data, it ***must be false***.

In 1633, this same Bellarmine (later made into a saint of the church) prosecuted Galileo in the Inquisition. Galileo was found "vehemently suspect of heresy" for precisely the reasons laid out in Bellarmine's original letter to Galileo. He was forced to publicly recant, his book laying out the reasons for believing the Copernican model was added along with the book of Copernicus to the list of banned books, and he was sentenced to live out his life under house arrest, praying all day for forgiveness. He died in 1642 a broken man, his prodigious and productive mind silenced by the active defenses of the locally dominant religious mythology for almost ten years.

I was fortunate enough to be teaching gravitation in the classroom on October 31, 1992, when Pope John Paul II (finally) *publicly apologized* for how the entire Galileo affair was handled. On Galileo's behalf, I accepted the apology, but of course I must also point out that *Bellarmine's argument is essentially correct*. The conclusions of modern science have, almost without exception, contradicted the assertions made in the holy scriptures not just of Christianity but of all faiths. They therefore stand as direct evidence that those scriptures are *not*, in fact, divinely inspired or perfect truth, at least where we can check them. While this does not *prove* that they are incorrect in other claims made elsewhere, it certainly and legitimately makes them less plausible.

12.2: Kepler's Laws

Galileo was not, in fact, the person who made the greatest contributions to the rejection of the Ptolemaic model as the first step towards first the (better) heliocentric Copernican model, then to the invention of physics and science as a systematic methodology for successively improving our beliefs about the Universe that does not depend on authority or scripture. He wasn't even one of the top two. Let's put him in the third position and count up to number one.

[284] http://lifesnow.com/bingo/ http://lifesnow.com/bingo/

The person in the second position (in my opinion, anyway) was *Tycho Brahe*[285] , a wealthy Danish nobleman who in 1571, upon the death of his father, established an observatory and laboratory equipped with the most modern of contemporary instrumentation in an abbey near his ancestral castle. He then proceeded to spend a substantial fraction of his life, including countless long Danish winter nights, *making and recording systematic observations of the night sky!*

His observations bore almost immediate fruit. In 1572 he observed a supernova in the constellation Cassiopeia. This one observation refuted a major tenet of Aristotelian and Church philosophy – that the Universe beyond the Moon's orbit was immutable. A new star had appeared where none was observed before. However, his most important contributions were immense tables of very precise measurements of the locations of objects visible in the night sky, over time. This was in no small part because his *own* hybrid model for a mixture of Copernican and Ptolemaic motion proved utterly incorrect.

If you are a wealthy nobleman with a hobby who is generating a huge pile of data but who also has no particular mathematical skill, what are you going to do? You hire a lab rat, a flunky, an *assistant* who can do the annoying and tedious work of analyzing your data while you continue to have the pleasure of accumulating still more. And as has been the case many a time, the servant exceeds the master. The number one philosopher who contributed to the Copernican revolution, more important than Brahe, Bruno, Galileo, or indeed any natural philosopher before Newton was Brahe's assistant, *Johannes Kepler*[286] .

Kepler was a brilliant young man who sought geometric order in the motions of the stars and planets. He was also a protestant living surrounded by Catholics in predominantly Catholic central Europe and was persecuted for his religious beliefs, which had a distinctly negative impact on his professional career. In 1600 he came to the attention of Tycho Brahe, who was building a new observatory near Prague. Brahe was impressed with the young man, and gave him access to his closely guarded data on the orbit of Mars and attempted to recruit him to work for him. Although he was was trying hard to be appointed as the mathematician of Archiduke Ferdinand, his religious and political affilations worked against him and he was forced to flee from Graz to Prague in 1601, where Brahe supported him for a full year until Brahe's untimely death (either from possibly deliberate mercury poisoning or a bladder that ruptured from enforced continence at a state banquet – it isn't clear which even today). With Brahe's support, Kepler was appointed an Imperial mathematician and "inherited" at least the use of Brahe's voluminous data. For the next eleven years he put it to very good use.

Although he was largely ignored by contemporaries Galileo and Descartes, Kepler's work laid, as we shall see, the foundation upon which one Isaac Newton built his physics. That foundation can be summarized in ***Kepler's Laws*** describing the motion of the orbiting objects of the solar system. They were *observational* laws, propounded on the basis of careful analysis of the Brahe data and further observations to verify them. Newton was able to *derive* trajectories that rather precisely agreed with Kepler's Laws on the basis of his physics and law of gravitation.

The laws themselves are surprisingly simple and geometric:

[285]Wikipedia: http://www.wikipedia.org/wiki/Tycho Brahe.
[286]Wikipedia: http://www.wikipedia.org/wiki/Johannes Kepler.

a) Planets move around the Sun in ***elliptical orbits*** with the Sun at one focus (see next section for a review of ellipses).

b) Planets sweep out ***equal areas in equal times*** as they orbit the Sun.

c) The ***mean radius*** of a planetary orbit (in particular, the semimajor axis of the ellipse) ***cubed*** is directly proportional to the ***period*** of the planetary orbit ***squared***, with the same constant of proportionality for all of the planets.

The first law can be proven directly from Newton's Law of Gravitation (although we will not prove it in this course, as the proof is mathematically involved). Instead we will content ourselves with the observation that a circular orbit is certainly consistent, and by using energy diagrams we will see that elliptical orbits are at least rather plausible. The second law will turn out to be equivalent to the ***conservation of angular momentum*** of the orbits, because gravitation is a ***central force*** and exerts no torque. The third, again, is difficult to formally prove for elliptical orbits but straightforward to verify for circular orbits.

Since most planets have *nearly* circular orbits, we will not go far astray by idealizing and restricting our analysis of orbits to the circular case. After all, not even elliptical orbits are precisely correct, because Kepler's results and Newton's demonstration *ignore the influence of the planets on each other* as they orbit the Sun, which constantly perturb even elliptical orbits so that they are at best a not-quite-constant approximation. The best one can do is directly and numerically integrate the equations of motion for the entire solar system (which can now be done to quite high precision) but even that eventually fails as small errors from ignored factors accumulate in time.

Nevertheless, the path from Ptolemy to Copernicus, Galileo and Kepler to Newton stands out as a great triumph in the intellectual and philosophical development of the human species. It is for that reason that we study it.

12.2.1: Ellipses and Conic Sections

The following is a short review of the properties of ellipses (and, to a lesser extent, the other conic sections). Recall that a conic section is the intersection of a plane with a right circular cone aligned with (say) the z-axis, where the intersecting plane can intercept at any value of z and parallel, perpendicular, or at an angle to the x-y plane.

A circle is the intersection of the cone with a plane parallel to the x-y plane. An ellipse is the intersection of the cone with a plane tipped at an angle *less* than the angle of the cone with the cone. A parabola is the intersection of the cone with a plane *at the same angle* as that of the cone. A hyperbola is the intersection of the cone with a plane tipped at a *greater* angle than that of the cone, so that it produces two disjoint curves and has *asymptotes*. An example of each is drawn in figure 12.2, the hyperbola for the special case where the intersecting plane is parallel to the z-axis.

Properly speaking, gravitational two-body orbits are conic sections: hyperbolas, parabolas, ellipses, or circles, not just ellipses per se. However, *bound planetary orbits* are elliptical, so we will concentrate on that.

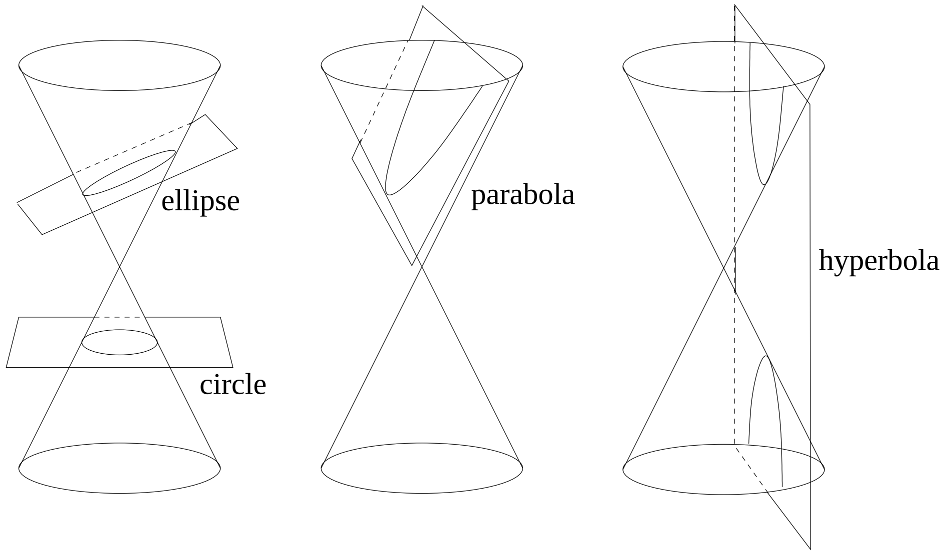


Figure 12.2: The various conic sections. Note that a circle is really just a special case of the ellipse.

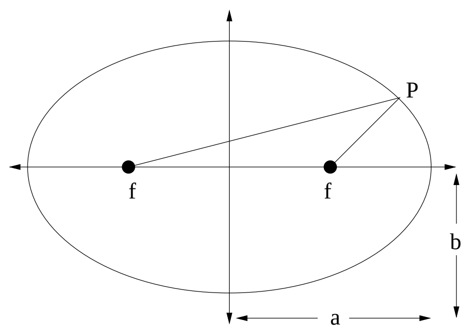


Figure 12.3

Figure 12.3 illustrates the general geometry of the ellipse in the x-y plane drawn such that its major axis is aligned with the x axis. In this simple case the equation of the ellipse can be written:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{12.20}$$

There are certain terms you should recall that describe the ellipse. The major axis is the longest "diameter", the one that contains both foci and the center of the ellipse. The minor axis is the shortest diameter and is at right angles to the major axis. The semimajor axis is the long-direction "radius" (half the major axis); the semiminor axis is the short-direction "radius" (half the minor axis).

In the equation and figure above, a is the semimajor axis and b is the semiminor axis.

Not all ellipses have major/minor axes that can be easily chosen to be x and y coordinates. Another general parameterization of an ellipse that is useful to us is a parametric cartesian representation:

$$x(t) = x_0 + a\cos(\omega t + \phi_x) \tag{12.21}$$
$$y(t) = y_0 + b\cos(\omega t + \phi_y) \tag{12.22}$$

This equation will describe *any* ellipse centered on (x_0, y_0) by varying ωt from 0 to 2π. Adjusting the phase angles ϕ_x and ϕ_y and amplitudes a and b vary the orientation and eccentricity of the ellipse from a straight line at arbitrary angle to a circle.

The ***foci*** of an ellipse are defined by the property that the sum of the distances from the foci to every point on an ellipse is a constant (so an ellipse can be drawn with a loop of string and two thumbtacks at the foci). If f is the distance of the foci from the origin, then the sum of the distances must be $2d = (f + a) + (a - f) = 2a$ (from the point $x = a$, $y = 0$. Also, $a^2 = f^2 + b^2$ (from the point $x = 0$, $y = b$). So $f = \sqrt{a^2 - b^2}$ where by convention $a \geq b$.

This is all you need to know (really more than you need to know) about ellipses in order to understand Kepler's First and Third Laws. The key things to understand are the meanings of the terms "focus of an ellipse" (because the Sun is located at one of the foci of an elliptical orbit) and "semimajor axis" as a measure of the "average radius" of a periodic elliptical orbit. As noted above, we will concentrate in *this* course on circular orbits because they are easy to solve for and understand, but in future, more advanced physics courses students will actually solve the equations of motion in 2 dimensions (the third being irrelevant) for planetary motion using Newton's Law of Gravitation as the force and prove that the solutions are parametrically described ellipses. In some versions of even *this* course, students might use a tool such as octave, mathematica, or matlab to solve the equations of motion numerically and graph the resulting orbits for a variety of initial conditions.

12.3: Newton's Law of Gravitation

In spite of the church's opposition, the early seventeenth century saw the formal development of the heliocentric hypothesis, supported by Kepler's empirical laws. Instrumentation improved, and the geometric methods involving *parallax* to determine distance produced a systematically improving picture of the solar system that was not only heliocentric but verified Kepler's Laws in detail for additional planetary bodies. The debate with the geocentric/ptolemaic model supporters continued, but in countries far away from Rome where its influence waned, a consensus was gradually forming that the geocentric hypothesis was incorrect. The observations of Brahe and Galileo and analysis of Kepler was compelling.

However, the *cause* of heliocentric motion was a mystery. There was clearly substantial geometry and order in the motion of the planets, although it was not precisely the geometry proposed by Plato and advanced by Aristotle and Ptolemaeus and others. This geometry was subtle, and best described within the confines of the new Analytic Geometry invented by Descartes[287] where ellipses (as we can see above) were not "just" conic sections or objects visualized in a solid geometry: They could be represented by *equations*.

Descartes was another advocate of the heliocentric theory, but when, in 1633, he heard that Galileo had been condemned for his advocacy of Copernicus and arguments against the

[287]Wikipedia: http://www.wikipedia.org/wiki/Rene Descartes. Descartes was another of the "renaissance man" polymaths of the age. He was brilliant and led a most interesting life, making contributions to mathematics (where "Cartesian Coordinates" are named in his honor), physics, and philosophy. He reportedly liked to sleep late, never rising before 11 a.m., and when an opportunity to become a court mathematician and tutor arose that forced him to change his habits and arise at 5 a.m. every day, he sickened and died (in 1650) a short while thereafter!

Ptolemaic geocentric model, he abruptly changed his mind about publishing a work to that effect! As noted above, these were dangerous times for freethinking philosophers who were literally forbidden by the rulers of the predominant religion under threat of torture and murder from speculating in ways that contradicted the scriptures of that religion. A powerful voice was thus silenced and the geocentric model persisted without any *open* challenge for fifty more years.

So things remained until one of the most brilliant and revered men of all time came along: Isaac Newton. Born on December 25, 1642, Newton was only 8 in 1650 when Descartes died, but he was taught Descartes' geometry at Cambridge (before it closed in the midst of a bout of the plague so that he was sent home for a while) and by the age of 24 had transformed it into a theory of "fluxions" – the first rudimentary description of calculus. Calculus, or the mathematics of related rates of change established on top of a coordinatized geometry, was the missing ingredient, the key piece needed to transform the *strictly geometric* observations of philosophers from Plato through Kepler into an *analytic description* of both the causes and effects of motion.

Even so, Newton worked *thirteen more years* producing and presenting advances in mathematics, optics, and alchemy before (in 1679), having recently completed a speculative theory of optics, he turned his attention wholly towards the problem of celestial mechanics and Kepler's Laws. In this he was reportedly inspired by the intuition that the force of *gravity* – the same force that makes the proverbial apple fall from the tree – was responsible for holding the moon in its orbit around the Earth.

Initially he corresponded heavily with *Robert Hooke*[288] , known to us through *Hooke's Law* in the text above, who had been appointed secretary of the brand new *Royal Society*[289] , the world's first "official" scientific organization, devoted to an eclectic mix of mathematics, philosophy, and the brand new "natural philosophy" (the correct and common terminology for "science" almost to the end of the nineteenth century). Hooke later claimed (quite possibly correctly) that he suggested the inverse-square force law to Newton, but what Hooke did not do that Newton did is to take the postulated inverse square force law, add to it a set of axioms (Newton's Laws) that *defined* force in a particular mathematical way, and then show that the equations of motion that followed from an inverse square force law, evaluated through the use of calculus, *completely predicted and explained Kepler's Laws and more* by means of explicit functional solutions built on top of Descartes' analytic geometry, where the "more" was the apparent non-elliptical orbits of other celestial bodies, notably comets.

It is difficult to properly explain how revolutionary, how world-shattering this combination of invention and discovery was. Initially it was communicated privately to the Royal Society itself in 1684; three years later it was formally published as the *Philosophiae Naturalis Principia Mathematica*[290] , or "The Mathematical Principles of Natural Philosophy". This book changed everything. It utterly destroyed, forever, any possibility that the geocentric hypothesis was correct. The reader must determine for themselves if it initiated the very process anticipated and feared by Robert Bellarmine – as the consequences of Newton's work unfolded, they have proven the Bible and all of the other religious mythologies and scriptures of the world *literally*

[288]Wikipedia: http://www.wikipedia.org/wiki/Robert Hooke.
[289]Wikipedia: http://www.wikipedia.org/wiki/Royal Society.
[290]Wikipedia: http://www.wikipedia.org/wiki/Philosophiae Naturalis Principia Mathematica.

false time and again.

As we have seen from a full semester of work with its core principles, Newton's Laws and a small set of actual force laws permit the nearly full description and prediction of virtually all everyday mechanical phenomena, and its *ideas* (in some cases extended far beyond what Newton originally anticipated) survive to some extent even in its eventual replacement, quantum mechanics. *Principia Mathematica* laid down a template for the *process* of scientific endeavor – a mix of accumulation and analysis of experimental data, formal axiomatic mathematics, and analytic reasoning leading to a detailed description of the visible Universe of ever-improving consistency. It was truly a *system of the world*, the basis of the ***scientific worldview***. It was a radically different worldview than the one based on faith, authority, and the threat of violence divine or mundane to any that dared challenge it that preceded it.

Let us take a look at the force law invented or discovered (as you please) by Newton and see how it works to explain Kepler's Laws, at least for simple cases we can readily solve without much calculus.

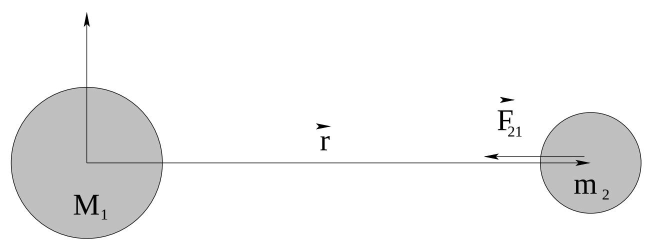


Figure 12.4

Here are Newton's axioms, the essential individual assumptions that are assembled compactly into the law of gravitation. Note that these assumptions were initially applied to objects like the Sun and the planets and moons that are spherically symmetric to a close approximation; the also apply to "particles" of mass or chunks of mass small enough to be treated as particles. Following along with figure 12.4 above:

a) The force of gravity is a ***two body force*** and does not change if three or more bodies are present.

b) The force of gravity is ***action at a distance*** and does not require the two objects to "touch" in order to act.

c) The force of gravity acts along (in the direction of) a line ***joining centers of spherically symmetric masses***, in this case along $\vec{r}$.

d) The force of gravity is ***attractive***.

e) The force of gravity is ***proportional to each mass***.

f) The force of gravity is ***inversely proportional to the distance between the centers of the masses***.

We will add to this list the assumption that one of the two masses is *much larger than the other* so that the center of mass and the center of coordinates can both be placed at the center of the larger mass. This is *not at all necessary* and proper treatments dating all the way back to Newton account for motion around a more general center of mass, but for us it will greatly simplify our pictures and treatments if we idealize in this way and in the case of systems like the Earth and the moon, or the Sun and the Earth, it isn't a terrible idealization. The Sun's mass is a thousand times larger than even that of Jupiter!

These axioms are rather prolix in words, but in the form of an *algebraic equation* they are rather *beautiful*:

$$\vec{F}_{21} = -\frac{G\ M_1\ m_2}{r^2}\hat{r} \tag{12.23}$$

where $G = 6.67 \times 10^{-11}$ N-m^2/kg^2 is the **universal gravitational constant**, the constant of proportionality that establishes the connections between all of the different *units* in question. Note that we continue to use the convention that $\vec{F}_{21}$ stands for the force acting on mass 2 due to mass 1; the force $\vec{F}_{12} = -\vec{F}_{21}$ both from Newton's third law and because the force is *attractive* for both masses.

Kepler's first law follows from solving Newton's laws and the equations of motion in three dimensions for this particular force law. Even though one dimension turns out to be irrelevant (the motion is strictly in a plane), even though the motion turns out to have two constants of the motion that permits it to be further simplified (e.g. the energy and the angular momentum) the actual solution of the resulting differential equations is a bit difficult and beyond the scope of this course. We will instead show that *circular* orbits are *special, tractable* solutions that easily satisfy Kepler's First and Third Laws, while Kepler's Second Law is a trivial consequence of conservation of angular momentum.

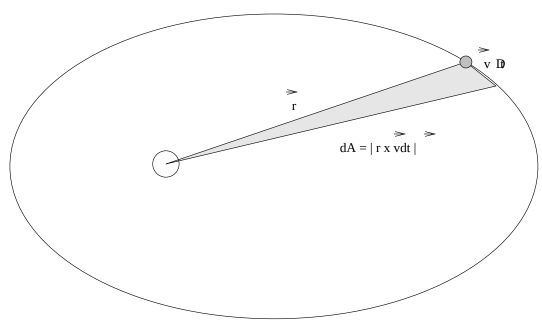


Figure 12.5: The area swept out in an elliptical orbit in time Δt is shaded in the ellipse above.

Let us begin with Kepler's Second Law, as it stands alone (the other two proofs are related). It is proven by observing that the force is ***radial*** along the line connecting the centers, and hence exerts ***no torque*** about a central pivot! That is:

$$\vec{\tau} = \vec{r} \times \left(-\frac{GM_1 m_2}{r^2}\hat{r}\right) = 0 = \frac{d\vec{L}}{dt}$$

so the angular momentum $\vec{L}$ of a planetary orbit is constant!

To show how this leads to Kepler's Second Law, we start by noting that the area enclosed by an parallelogram formed out of two vectors is the magnitude of the the cross product of those vectors. Hence the area in the shaded triangle in figure 12.5 is half of that:

$$dA = \frac{1}{2}|\vec{r} \times \vec{v}\, dt| = \frac{1}{2}|\vec{r}||\vec{v}\, dt| \sin\theta \tag{12.24}$$

$$= \frac{1}{2m}|\vec{r} \times m\vec{v}\, dt| \tag{12.25}$$

If we divide the Δt over to the other side we get the area per unit time being swept out by the orbit:

$$\frac{dA}{dt} = \frac{1}{2m}|\vec{r} \times \vec{p}| = \frac{1}{2m}|\vec{L}| = \text{a constant} \tag{12.26}$$

because angular momentum is conserved for a central force as shown above. Kepler's second law is therefore proved for this force, and indeed, will be true for *any* radial/central force law, even one that is *not* e.g. an inverse square law!

That was pretty easy! Let's reiterate the point of this demonstration:

Kepler's Second Law is equivalent to the Law of Conservation of Angular Momentum and is true for any central force (not just gravitation)!

The proofs of Kepler's First and the Third laws *for circular orbits* rely on a common algebraic argument, so we group them together. They key formula is, as one might expect given our knowledge (from the very first chapter of this textbook) that ***if*** an orbiting mass moves in a *circular* orbit, ***then*** the gravitational force has to be equal to the mass times the centripetal acceleration:

$$\frac{G\, M_s m_p}{r^2} = m_p a_r = m_p \frac{v^2}{r} \tag{12.27}$$

where M_s is the mass of the central attracting body (which we implicitly assume is much larger than the mass of the orbiting body so that its center of mass is more or less at the center of mass of the system), m_p is the mass of the planet, v is its speed in its circular orbit of radius r. This situation is illustrated in figure 12.6.

This equation in and of itself "proves" that Newton's Laws plus Newton's Law of Gravitation have a solution consisting of a circular orbit, where a circle is a special case of an ellipse. This proof isn't very exciting, however, as *any* attractive radial force law we might attempt would have a similarly consistent circular solution. The kinematic radial acceleration of a particle moving in uniform circular motion is independent of the particular force law that produces it!

What is a lot more interesting is the demonstration that the circular orbit satisfies Kepler's *Third* Law, as this law quite specifically defines the relationship between the radius of the orbit and its period. We can easily see that only *one* radial force law will lead to consistency with the observational data for circular orbits.

We start by cancelling the mass of the planet and one of the factors of r:

$$v^2 = \frac{G\, M_s}{r} \tag{12.28}$$

But, v is related to r and the period T by:

$$v = \frac{2\pi r}{T} \tag{12.29}$$

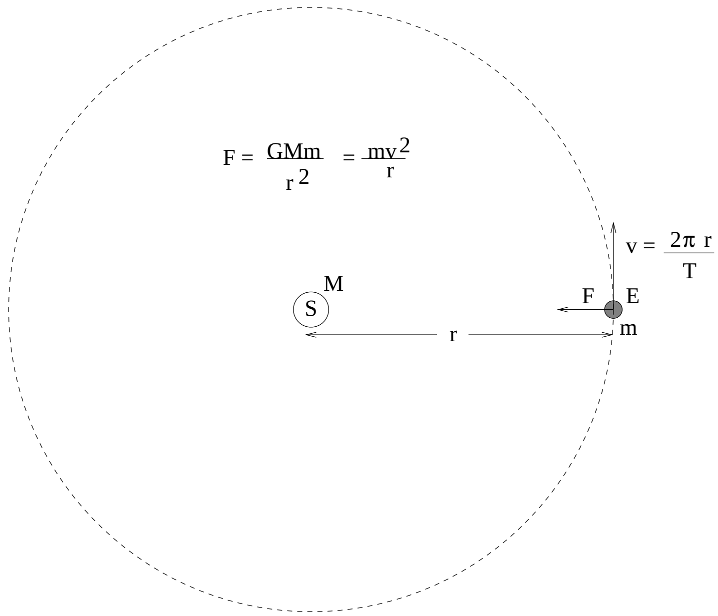


Figure 12.6: The geometry used to prove Kepler's and Third Laws for a circular (approximately) orbit like that of the Earth around the Sun.

so that

$$v^2 = \frac{4\pi^2 r^2}{T^2} = \frac{G\ M_s}{r} \tag{12.30}$$

Finally, we isolate the powers of r:

$$r^3 = \left(\frac{G\ M_s}{4\pi^2}\right) T^2 \tag{12.31}$$

and Kepler's third law is proved for circular orbits.

Since there is nothing unique about circular orbits and (empirically, at least) *all* closed elliptical orbits around the *same* central attracting body are found to have the *same constant of proportionality*, we have both proven that Newton's Law of Gravitation has circular solutions that satisfy Kepler's Third Law *and* we have obtained an algebraic expression for the *universal* constant of proportionality, valid for all of the planets in the solar system! We can then write the law more compactly:

$$R_{sm}^3 = \left(\frac{G\ M_s}{4\pi^2}\right) T^2 \tag{12.32}$$

where now R_{sm} is the semimajor axis of the elliptical orbit, which happens to be r for a circular orbit.

Note well that this constant is easily measured! In fact we can evaluate it from our knowledge of the semimajor axis of Earth's nearly circular orbit – $R_E \approx 1.5 \times 10^{11}$ meters (150 million kilometers) plus our knowledge of its period – $T = 3.153 \times 10^7$ seconds (1 year, in seconds). These two numbers are well worth remembering – the first is called an ***astronomical unit*** and is one of the fundamental lengths upon which our knowledge of the distances to the nearer stars is based; the second physicists tend to remember as "ten million times π seconds per year" because that is accurate to well within one percent and easier to remember than 3.153.

Combining the two we get:

$$\left(\frac{G\,M_s}{4\pi^2}\right) = \frac{T^2}{R_{sm}^3} = \frac{\pi^2 \times 10^{14}}{3.375 \times 10^{33}} \approx 3 \times 10^{-19} \tag{12.33}$$

where we used *another* physics geek cheat: $\pi^2 \approx 10$, and then approximated $10/3.375 \approx 3$ as well. That way we can get an answer, good to within a couple of percent, without using a calculator or looking anything up!

Note well! If only we knew G, we'd know the mass of the Sun! If we use the same logic to determine the *same* constant for objects orbiting the *Earth* (where we might use the semimajor axis of the moon's orbit, 384,000 kilometers, and the period of the moon's orbit, 27.3 days, to get $GM_E/4\pi^2$) we would also be able to determine the mass of the Earth!

Of course we *do* know G *now*, but when Newton proposed his theory, it wasn't so easy to figure out! This is because *gravitation is the weakest of the forces of nature, by far!* It is so weak that it is remarkably difficult to measure the direct gravitational force between two objects of known masses separated by a known distance in the laboratory, so that all of the quantities in Newton's Law of Gravitation were measured *but* G.

In fact, it took over a century for Henry Cavendish[291] to build a clever apparatus that was sufficiently sensitive that it could measure G from these three known quantities. This experiment was said to "weigh the Earth" not because it actually did so – far from it – but because once G was known experiments that had long since been done instantly gave us the mass of the Sun, the mass of the Earth, the mass of Jupiter and Saturn and Mars (any planet where we can remotely observe the semimajor axis and period of a moon) and much more.

These in turn gave us some serious conundrums! The Sun turns out to be 1.4 *million kilometers* in diameter, and to have a mass of 2×10^{30} kilograms! With a surface temperature of some 6000 K, what mechanism keeps it so hot? Any sort of chemical fire would soon burn out!

Laboratory experiments plus astronomical observations based on the use of parallax with the entire diameter of the Earth's orbit used as a triangle base and with exquisitely sensitive measurements of the angles between the lines of sight to the nearer stars (which allowed us to determine the distance to these stars) all analyzed by means of Newton's Laws (including gravitation), allowed astronomers to rapidly infer a startling series of facts about the Solar system, our local galaxy (the Milky Way), and the Earth.

Not only was the *geocentric* hypothesis wrong, so was the *heliocentric* hypothesis. The Earth turned out to be a mostly unremarkable planet, a relatively small one of a rather large number orbiting an entirely unremarkable star that itself was orbiting in a huge collection of stars, that was only one of a truly staggering number of similar collections of stars, where every new generation of telescopes revealed still more of everything, still further away. At the moment, there appear to be on the order of a hundred *billion* galaxies, containing somewhere in the ballpark of 10^{23} stars, in the visible Universe, which is (allowing for its original inflation, 13.7 billion years ago) around 46 billion light years in radius. At least one method of estimation has claimed to establish a radius around twice this large as a *lower bound* for its size (so that all of these estimates are probably low by an order of magnitude) – and there is no upper bound.

[291] Wikipedia: http://www.wikipedia.org/wiki/Cavendish Experiment.

Exoplanets are being discovered at a rate that suggests that planetary systems around those stars are *common*, not rare (especially so given that we can only "see" or infer the existence of extremely large planets so far – we would find it almost impossible to detect a planet as small as the Earth). Bruno's original assertion that the Universe is infinite, contains and infinite number of stars, with an infinite number of planets, an infinite number of which have some sort of intelligent or otherwise *life* may be impossible to verify or refute, but infinite or not the Universe is *enormous* compared to the scale of the Solar system, which is *huge* compared to the scale of the Earth, and contains many, many stars with many, many planetary systems.

In fact, the only thing about the Earth that is remarkable may turn out to be – us!

12.4: The Gravitational Field

As noted above, Newton proposed the gravitational force as the *cause* of the observed orbital motions of the celestial objects. However, this force was *action at a distance* – it exists between two objects that are not touching and that indeed are separated by *nothing*: a vacuum! What then, causes the gravitational force itself? Let us suggest that there must be *something* that is produced by one planet acting as a *source* that is present at the location of the other planet that is the proximate cause of the force that planet experiences. We define the ***gravitational field*** to be this ***cause*** of the gravitational force, the thing that is present at all points in space surrounding a mass *whether or not some other mass is present there to be acted on!*

We define the gravitational field conveniently to be the force per unit mass, a quantity that has the units of *acceleration*:

$$\vec{\boldsymbol{g}}(\vec{\boldsymbol{r}}) = -\frac{G\ M}{r^2}\hat{\boldsymbol{r}} = \frac{\vec{\boldsymbol{F}}}{m} \tag{12.34}$$

The magnitude of the gravitational field at the surface of the earth is thus:

$$g = g(R_E) = \frac{F}{m} = \frac{G\ M_E}{R_E^2} \tag{12.35}$$

and we see that the quantity that we have been calling the gravitational *acceleration* is in fact more properly called the near-Earth gravitational *field*.

This is a very useful equation. It can be used to find any one of g, R_E, M_E, or G, from a knowledge of any of the other three, depending on which ones you think you know best. g is easy; students typically measure g in physics labs at some point or another several different ways! R_E is actually also easy to measure independently and some classical methods were used to do so *long* before Columbus.

M_E, however is hard! This is because it always appears in the company of G, so that knowing g and R_E only gives you their product. This turns out to be the case nearly everywhere – any ordinary measurement you might make turns out to tell you GM_E together, not either one separately.

What about G?

To measure G in the laboratory, one needs a very sensitive apparatus for measuring forces. Since we know already that G is on the order of 10^{-10} N-m^2/kg^2, we can see that gravitational forces between kilogram-scale masses separated by ten centimeters or so are on the order of a few *billionths* of a Newton.

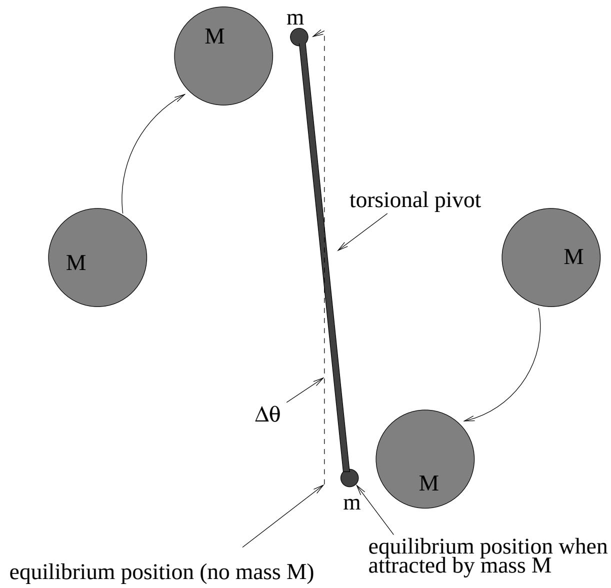


Figure 12.7: The apparatus associated with the ***Cavendish experiment***, which established the first accurate estimates for G and thereby "weighed the Earth", the Sun, and many of the other objects we could see in the sky.

Henry Cavendish made the first direct measurement of G using a torsional pendulum – basically a barbell suspended by a very thin, strong thread – and some really massive balls whose relative position could be smoothly adjusted to bring them closer to and farter from the barbell balls. As you can imagine, it takes very little torque to twist a long thread from its equilibrium angle to a new one, so this apparatus has – when utilized by someone with a great deal of patience, using a light source and a mirror to further amplify the resolution of the twist angle – proven to be sufficiently sensitive to measure the tiny forces required to determine G, even to some reasonable precision.

Using this apparatus, he was able to find G and hence to "weigh the earth" (find M_E). By measuring $\Delta\theta$ as a function of the distance r measured between the centers of the balls, and calibrating the torsional response of the string using known forces, he managed to get 6.754 (vs 6.673 currently accepted) $\times 10^{-11}$ N-m^2/kg^2. This is within just about one percent. Not bad!

12.4.1: Spheres, Shells, General Mass Distributions

So far, our empirically founded expression for gravitational force (and by inheritance, field) applies only to ***spherically symmetric mass distributions*** – planets and stars, which are generally almost perfectly round *because* of the gravitational field – or particles small enough that they can be treated like spheres. Our pathway towards the gravitational field of more general distributions of mass starts by formulating the field of a single point-like chunk of mass in such a distribution:

$$d\vec{g} = -\frac{G\,dm_0}{|\vec{r} - \vec{r}_0|^3}(\vec{r} - \vec{r}_0) \tag{12.36}$$

This equation can be integrated as usual over an arbitrary mass distribution using the usual connection: The mass of each chunk is the mass per unit volume times the volume of the chunk, or $dm = \rho dV_0$.

$$\vec{g} = -\int \frac{G\,\rho\,dV_0}{|\vec{r} - \vec{r}_0|^3}(\vec{r} - \vec{r}_0) \tag{12.37}$$

where for example $dV_0 = dx_0 dy_0 dz_0$ (Cartesian) or $dV_0 = r_0^2 \sin(\theta_0) d\theta_0 d\phi_0 dr_0$ (Sphereical Polar) etc. This integral is not always easy, but it can generally be done very accurately, if necessary numerically. In simple cases we can actually do the calculus and evaluate the integral.

In *this* part of *this* course, we will avoid doing the integral, although we will tackle many examples of doing it in simple cases next semester. We will content ourselves with learning the following **True Facts** about the gravitational field:

- The gravitational field produced by a (thin) spherically symmetric shell of mass ΔM vanishes inside the shell.
- The gravitational field produced by this same shell equals the usual

  $$\vec{g}(\vec{r}) = -\frac{G\Delta M}{r^2}\hat{r} \tag{12.38}$$

  outside of the shell. As a consequence the field outside of any spherically symmetric distribution of mass is just

  $$\vec{g}(\vec{r}) = -\frac{G\Delta M}{r^2}\hat{r} \tag{12.39}$$

These two results can be proven by direct integration or by using Gauss's Law for the gravitational field (using methodology developed next semester for the electrostatic field). The latter is so easy that it is hardly worth the time to learn the former for this special case.

Note well the most important consequence for our purposes in the homework of this rule is that when we descend a tunnel into a uniformly dense planet, the gravity will *diminish* as we are only pulled down by the mass *inside* our radius. This means that the gravitational field we experience is:

$$\vec{g}(\vec{r}) = -\frac{G\Delta M(r)}{r^2}\hat{r} \tag{12.40}$$

where $M(r) = \rho 4\pi r^3/3$ for a uniform density, something more complicated in cases where the density itself changes with r. You will use this expression in several homework problems.

12.5: Gravitational Potential Energy

If you examine figure 12.8 above, and note that the force is always "down" along $\vec{r}$, it is easy to conclude that gravity must be a conservative force. Gravity produced by some (spherically symmetric or point-like) mass does work on another mass only when that mass is moved in or

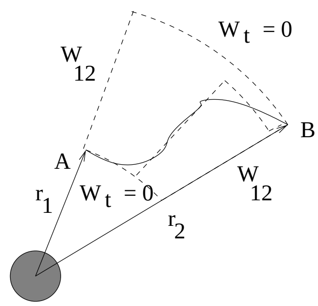


Figure 12.8: A crude illustration of how one can show the gravitational force to be conservative (so that the work done by the force is independent of the path taken between two points), permitting the evaluation of a potential energy function.

out along $\vec{r}$ connecting them; moving at right angles to this along a surface of constant radius r involves no gravitational work. Any path between two points near the source can be broken up into approximating segments parallel to $\vec{r}$ and perpendicular to $\vec{r}$ at each point, and one can make the approximation as good as you like by choosing small enough segments.

This permits us to easily compute the gravitational potential energy as the negative work done moving a mass m from a reference position $\vec{r}_0$ to a final position $\vec{r}$:

$$U(r) = -\int_{r_0}^{r} \vec{F} \cdot d\vec{r} \tag{12.41}$$

$$= -\int_{r_0}^{r} -\frac{GMm}{r^2} dr \tag{12.42}$$

$$= -\left(\frac{GMm}{r} - \frac{GMm}{r_0}\right) \tag{12.43}$$

$$= -\frac{GMm}{r} + \frac{GMm}{r_0} \tag{12.44}$$

Note that the potential energy function depends only on the *scalar magnitude* of $\vec{r}_0$ and $\vec{r}$, and that r_0 is in the end the radius of an arbitrary point where we define the potential energy to be zero.

By convention, unless there is a good reason to choose otherwise, we require the zero of the gravitational potential energy function to be at $r_0 = \infty$. Thus:

$$U(r) = -\frac{GMm}{r} \tag{12.45}$$

Note that since energy in some sense is more fundamental than force (the latter is the negative derivative of the former) we could just as easily have learned Newton's Law of Gravitation directly as this *scalar* potential energy function and then evaluated the force by taking its negative gradient (multidimensional derivative).

The most important thing to note about this function is that it is *always negative*. Recall that the force points in the direction that the potential energy *decreases most strongly* in. Since $U(r)$ is negative and gets larger in magnitude for smaller r, gravitation (correctly) points *down* to smaller r where the potential energy is "smaller" (more negative).

The potential energy function will be very useful to us when we wish to consider things like escape velocity/energy, killer asteroids, energy diagrams, and orbits. Let's start with energy diagrams and orbits.

12.6: Energy Diagrams and Orbits

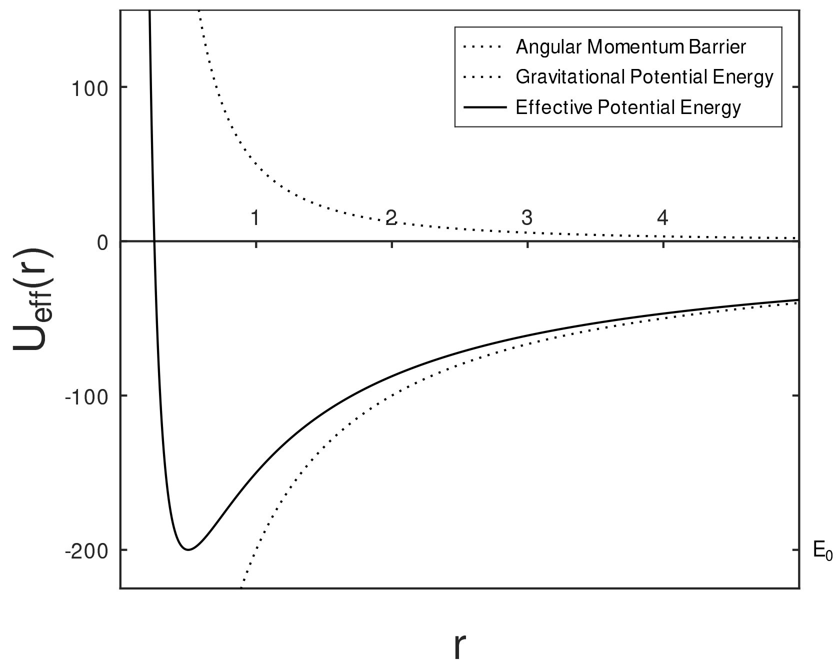


Figure 12.9: A plot of the *effective potential energy*: the sum of the true radial potential energy and the *rotational kinetic energy* of the orbiting object.

Let's write the total energy of a particle moving in a gravitational field in a clever way that isolates the ***radial kinetic energy*** and transforms the rest of the kinetic energy (arising from the component of $\vec{v}$ perpendicular to $\vec{r}$) into angular momentum form:

$$
\begin{aligned}
E_{\rm tot} &= \frac{1}{2}mv^2 - \frac{GMm}{r} \\
&= \frac{1}{2}mv_r^2 + \frac{1}{2}mv_\perp^2 - \frac{GMm}{r} = \frac{1}{2}mv_r^2 + \left\{\frac{mr^2}{mr^2} \times \frac{1}{2}mv_\perp^2\right\} - \frac{GMm}{r} \\
&= \frac{1}{2}mv_r^2 + \frac{1}{2mr^2}(mv_\perp r)^2 - \frac{GMm}{r} \\
&= \frac{1}{2}mv_r^2 + \left(\frac{L^2}{2mr^2} - \frac{GMm}{r}\right)
\end{aligned}
$$

where we multiplied by a clever form of 1 to transform the transverse kinetic energy term into rotational form using

$$L = mv_\perp r$$

(a ***constant of the motion*** for central forces like gravitation).

The angular/perpendicular part of the kinetic energy:

$$K_{\rm rot} = \frac{1}{2} m v_\perp^2 = \frac{L^2}{2mr^2} = \frac{L^2}{2I} \tag{12.46}$$

inside the parentheses above is called the ***angular momentum barrier***. It has units of energy, and (because L and m are both constants for any given orbit) *is a simple function of* r*!*

This motivates us to define the sum of the angular momentum barrier and the true gravitational potential energy (the quantity in parentheses) to be the ***effective radial potential energy function***:

$$U_{\rm eff}(r) = \frac{L^2}{2mr^2} - \frac{GMm}{r} \tag{12.47}$$

This is plotted in figure 12.9 where indeed it *looks like some of the potential energy functions we studied in Week 3!* When we add it to the radial kinetic energy, we get:

$$E_{\rm tot} = \frac{1}{2} m v_r^2 + U_{\rm eff}(r) \tag{12.48}$$

This looks *exactly like energy conservation with a one-dimensional potential energy and one dimensional kinetic energy!* We can, in fact, use it to draw and analyze *one-dimensional energy diagrams in the radial coordinate only* almost exactly the same way we learned in Week 3, as long as we remember that the kinetic energy in question is only the *radial component* of the total kinetic energy, describing how fast the orbiting particle is *approaching or receding from* the primary attractor!

Let's see how this works, and in the process learn a few very important thing about orbits and their classification! Let's take the graph of $U_{\rm eff}(r)$ above (which has units of *energy* on the ordinate) and decorate it with examples of possible total mechanical energies, which are of course constant for all r where they are physical.

As always, we need to recognize that the radial kinetic energy is given by the *difference*:

$$K_r = \frac{1}{2} m v_r^2 = E_{\rm tot} - U_{\rm eff} \geq 0 \tag{12.49}$$

and must, as indicated, be greater than or equal to zero in order for v_r to be real! As always, we identify the points where $v_r = 0$ to be *turning points*, after a fashion, but they won't be *quite* the same as true one-dimensional potential energy function turning points. At points where $K_r = 0 \Rightarrow v_r = 0$, it doesn't mean that the particle isn't moving – it still has *angular* kinetic energy if $L \neq 0$! It just means that the orbiting particle has reached either a *minimum or maximum distance* from the central attractor.

Those points are given special names of their own in "orbitspeak". The minimum distance of an orbit from the attractor is called the ***perihelion***. The maximum distance of an orbit from the attractor (if there is one – two kinds of orbits do not!) is called the ***aphelion***. These are the two ***apsides*** of the orbit[292] .

With all of this said, let's take a look, in figure **??**:

We can determine lots of interesting things from this diagram. It was generated for a fixed angular momentum $\vec{L} \neq 0$, and four possible total energies are drawn on it as straight horizontal lines (since total energy is conserved by the *conservative* Newtonian gravitation force). These orbits have the following characteristics and names:

[292] Wikipedia: http://www.wikipedia.org/wiki/Apsis.

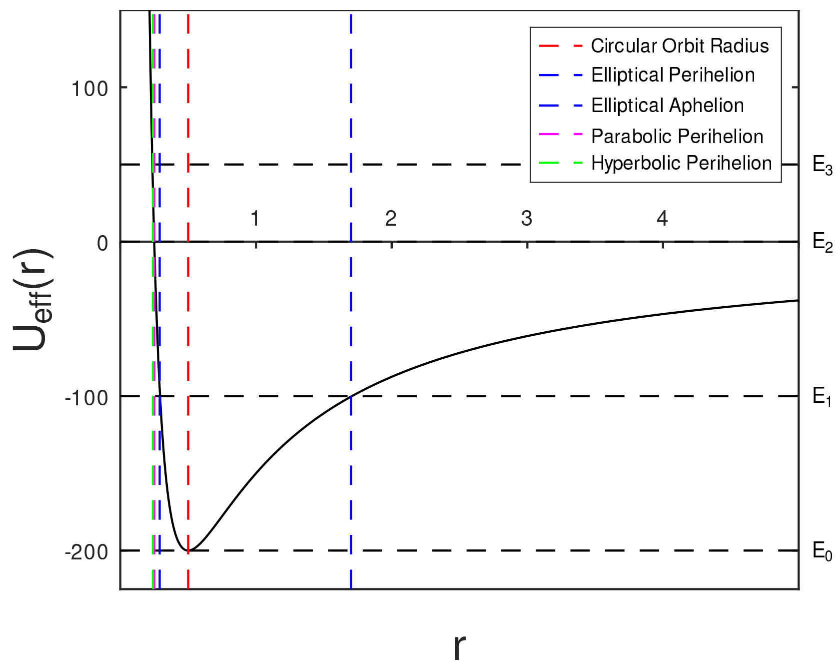


Figure 12.10: A radial total energy diagram illustrating the four distinct named orbits in terms of their total energy: E_0 is a ***circular*** orbit. E_1 is an ***elliptical*** orbit. E_2 is a ***parabolic*** orbit. E_3 is a ***hyperbolic*** orbit. Note that all of these orbits are *conic sections*, and that the classical elliptic orbits have two apsides (perigree and apogee) along the major axis of the ellipse.

a) $E_{\rm tot} = U_{\rm eff,min}$. This is a **circular** orbit. This is a special case of an elliptical orbit, but deserves special mention. A circular orbit has ***only one allowed radius***, one with ***zero radial kinetic energy!*** All of the kinetic energy is in the rotational kinetic energy of the circular orbit! Perigee (closest distance) and Apogee (farthest distance) are one!

b) $E_{\rm tot} < 0$. This is generally an **elliptical** orbit (consistent with Kepler's First Law). There is a true perigee/perihelion and a true apogee/aphelion on the semimajor axis of the ellipse.

c) $E_{\rm tot} = 0$. This is a **parabolic** orbit. This orbit defines **escape velocity** as we shall see in the next topic. It has a perihelion, but *no* aphelion – it has enough energy (barely) to ***make it all the way to infinity without ever quite ceasing to get still farther away***.

d) $E_{\rm tot} > 0$. This is a **hyperbolic** orbit. It has a perigee, but again no apogee – it has enough energy to get "to infinity and beyond" while still moving at a pretty good clip when it gets there.

Note well that in cosmological dynamics, "infinity" is really shorthand for "any distance that is so large that the gravitational field out there is negligible compared to that of everything *else* with gravitational fields, which is – everything else. By the time one is several hundred Earth radii out your gravitational potential energy isn't actually *zero*, but it is by then much less than 1% of the value it had on the Earth's surface. And this is still a *short* distance as far as the solar system goes, let alone the galaxy!

Note well that all of the orbits are **conic sections**. This interesting geometric connection between $1/r^2$ forces and conic section orbits was a tremendous motivation for important mathematical work two or three hundred years ago, when computers were unknown and a good deal of mathematics was worked out using geometry instead of algebra!

Next, let's look at a very important question. Imagine that you are down here on the surface of the Earth[293], looking up at the night sky. You see a satellite, you see the moon, you see the stars. You wonder: How much would it cost for me to go visit! And by cost, I don't mean in money (although that is certainly directly relevant as well).

I mean, cost in ***energy***[294]!

12.7: Escape Velocity, Escape Energy

As we noted in the previous section, a particle has "escape energy" if and only if its total energy is greater than or equal to zero, provided that we set the zero of potential energy at infinity in the first place. We define the ***escape velocity*** (a misnomer!) of the particle as the minimum ***speed*** (!) that it must have to escape from its current gravitational field – typically that of a moon, or planet, or star.

Suppose you are on the surface of a planet. What the heck, suppose you are on the surface of *Earth*[295]. You would like to do things like find your total mechanical energy (as you stand there), figure out how much kinetic energy you would have to add in the form of work to get to near Earth orbit, or perhaps to places still farther away – to the nearest *star*. Let's address these one at a time.

Durham, NC – where *I'm* sitting as I type this – is at 36° north latitude. This means that my personal distance from the axis of rotation is *about*:

$$R_{\rm rgb} = R_e \cos 36^\circ = 6400 \times 0.81 \approx 5200 \text{ kilometers}$$

The circumference at this latitude is 2π times this, or:

$$C_{\rm rgb} = 2\pi R_{\rm rgb} = 32600 \text{ kilometers}$$

(or 20300 miles). My *speed* as I sit still in my chair is, therefore:

Thus:

$$E_{\rm tot} = 0 = \frac{1}{2} m v_{\rm escape}^2 - \frac{GMm}{r} \tag{12.50}$$

so that

$$v_{\rm escape} = \sqrt{\frac{2GM}{r}} = \sqrt{2gr} \tag{12.51}$$

where in the last form $g = \frac{GM}{r^2}$ (the magnitude of the gravitational field – see next item).

To escape from the ***Earth's surface***, one needs to start with a speed of:

$$v_{\rm escape} = \sqrt{\frac{2GM_E}{R_E}} = \sqrt{2gR_E} = 11.2 \quad \text{km/sec} \tag{12.52}$$

[293] This shouldn't take much imagination. Hardly any, actually.

[294] Which generally has to be *paid for* with money, which is, after all, condensed work.

[295] As I said, this should hardly require any imagination at all...

Note: Recall the form derived by equating Newton's Law of Gravitation and mv^2/r in an earlier section for the velocity of a mass m in a circular orbit around a larger mass M:

$$v_{\rm circ}^2 = \frac{GM}{r} \tag{12.53}$$

from which we see that $v_{\rm escape} = \sqrt{2} v_{\rm circ}$.)

It is often interesting to contemplate this reasoning in reverse. If we drop a rock onto the earth from a state of rest "far away" (much farther than the radius of the earth, far enough away to be considered "infinity"), it will REACH the earth with escape (kinetic) energy and a total energy close to zero. Since the earth is likely to be much larger than the rock, it will undergo an *inelastic* collision and release nearly ***all its kinetic energy as heat***. If the rock is small, this is not necessarily a problem. If it is large – say, 1 km and up – it releases a *lot* of energy.

Example 12.7.1: How to Cause an Extinction Event

How much energy? Time to do an estimate, and in the process become just a tiny bit scared of a very, very unlikely event that could conceivably cause the extinction of *us*.

Let's take a "typical" rocky asteroid that might at any time decide to "drop in" for a one-way visit. While the asteroid might well have any shape – that of a potato, or *pikachu*[296] , we'll follow the usual lazy physicist route and assume that it is a simple spherical ball of rock with a radius r. In this case we can estimate its total mass as a function of its size as:

$$M = \frac{4\pi\rho}{3} r^3 \tag{12.54}$$

Of course, now we need to estimate its density, ρ. Here it helps to know two numbers: The density of water, or ice, is around 10^3 kg/m^3 (a metric ton per cubic meter), and the ***specific gravity or rock*** is highly variable, but in the ballpark of 2 to 10 (depending on how much of what kinds of metals the rock might contain, for example), say around 5.

If we then let $r \approx 1000$ meters (a bit over a mile in diameter), this works out to $M \approx 1.67 \times 10^{12}$ kg, or around 2 billion metric tons of rock, about the mass of a small mountain.

This mass will land on earth with *escape velocity*, 11.2 km/sec, if it falls in "from rest" from far away. Or more, of course – it may have started with velocity and energy from some other source – this is pretty much a minimum. As an exercise, compute the number of Joules this collision would release to toast the dinosaurs – or us! As a further exercise, convert the answer to "tons of TNT" (a unit often used to describe nuclear-grade explosions – the original nuclear fission bombs had an explosive power of around 20,000 tons of TNT, and the largest nuclear fusion bombs built during the height of the cold war had an explosive power on the order of 1 to 15 million tons of TNT.

The conversion factor is 4.184 gigajoules per ton of TNT. You can easily do this by hand, although the internet now boasts of calculators that will do the entire conversion for you. I get ballpark of ***ten to the twentieth*** joules or 25 gigatons – that is ***billions of tons*** – of TNT. In contrast, wikipedia currently lists the combined explosive power of all of the world's 30,000

[296]Wikipedia: http://www.wikipedia.org/wiki/Pikachu. If you don't already know, don't ask...

or so extant nuclear weapons to be around 5 gigatons. The explosion of Tambora (see last chapter) was estimated to be around 1 gigaton. The asteroid that might have caused the K-T extinction event that ended the Cretaceous and wiped out the dinosaurs and created the 180 kilometer in diameter ***Chicxulub crater***[297] had a diameter estimated at around 10 km and would have released around 1000 times as much energy, between 25 and 100 ***teratons*** of TNT, the equivalent of some 25,000 Tambora's happening all at once.

Such impacts are geologically rare, but obviously can have enormous effects on the climate and environment. On a smaller scale, they are one very good reason to oppose the military exploitation of space – it is all too easy to attack any point on Earth by dropping rocks on it, where the asteroid belt could provide a virtually unlimited supply of rocks.

12.8: The Tide

Waaaay back in Week/Chapter 5 we learned (proved!) that we could pretend that the force of gravity acting on a rigid body could be computed by pretending at all of the mass was concentrated at the center of mass, which *in the context of contant/uniform "near-Earth" gravity* could be considered the *center of gravity* as well.

This worked for *both* the total force (which was proven in Week/Chapter 4) *and* for the ***torque*** exerted on the mass due to gravitation. Although the proof wasn't generated for full three dimensional torques and pointlike pivots, it is still valid for them as long as the resultant torque is directed perpendicular to plane containing the pivot point and the center of mass/gravity. Basically:

> **For uniform gravitational fields, the center of mass and the center of gravity are the same**

This is *not* the case for non-uniform Newtonian gravitation! Gravitation in this case gets *weaker* the farther one gets from the center of the attracting body, e.g. the Earth or the Sun or the Moon, and worse, the *angle* at which the force acts varies as one sweeps across an extended object as well! This variation with r has profound consequences in the context of orbital motion, especially for objects that are large enough that the variation in ***both magnitude and direction*** of the gravitational field across the spatial extent of the object or system are *significant* relative to the magnitude and direction of the field at the center of mass.

The primary consequences of this variation are collectively referred to as "the ***tide***". It's not easy to reduce the meaning of this word to be quite identical to any of the forces or pseudoforces we've looked at so far – the easiest way to understand it will be to worth through it in some detail in a few "simple" cases. Our immediate conclusion after doing this is that in general tidal "forces" are quite complex and highly non-trivial, especially in the specific context of the Sun-Earth-Moon triple system that is relevant to us here on Earth, even if we ignore the (fortunately much smaller contributions) from the *rest* of the planets in the solar system. However, this tidal force is behind *actual* oceanic tide that we observe in sufficiently large bodies of water and is behind a number of other readily observable phenomena and hence is worth including even in an introductory physics textbook.

[297] Wikipedia: http://www.wikipedia.org/wiki/Chicxulum Crater.

The easiest way to learn about the tidal "force" (really a mix of pseudoforce and the $1/r^2$ variation of the gravitational field) is to look at a couple of comparatively simple "one dimensional" cases. In all cases tidal forces arise due to the *difference in magnitude and direction* of the gravitational field at different points on an extended object, plus contributions from pseudoforces in the accelerated frame of the object.

We'll start with a falling "dumbbell" – two identical masses at the ends of a massless rod, aligned with the r-varying field of a planetary attractor.

Example 12.8.1: A Freely Falling, Vertically Aligned Dumbell

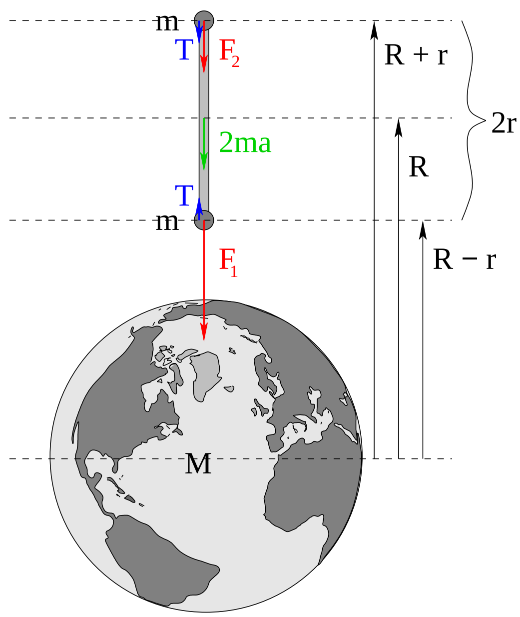


Figure 12.11: A massless rod of length $2r$ with identical point masses m located at the ends is located a distance $R > r$ away from a planetary center and aligned with the direction of the vector $\vec{R}$ to the center of mass of the rod/mass dumbbell.

Because everything is effectively one dimensional, it is easy to compute the total force acting on the dumbbell pictured in figure 12.11 (with "up" positive):

$$F_{v,\text{tot}} = -\frac{GMm}{(R+r)^2} - \frac{GMm}{(R-r)^2} \tag{12.55}$$

For very large $r \sim R$, there isn't a lot more that we can do, but in general any *real* problem involving the tide is likely enough to involve systems whose size is much smaller than (for example) the Earth.

For that reason, let's assume that $r \ll R$ and do a *binomial expansion* to solve for the leading order (in r/R) behavior of the total gravitational force:

$$\begin{aligned} F_{\rm tot} &= -\frac{GMm}{(R+r)^2} - \frac{GMm}{(R-r)^2} && (12.56) \\ &= -\frac{GMm}{R^2}\left\{(1+r/R)^{-2} + (1-r/R)^{-2}\right\} && (12.57) \\ &= -\frac{GMm}{R^2}\left\{(1 - 2\frac{r}{R} + 3\frac{r^2}{R^2} + ...) + (1 + 2\frac{r}{R} + 3\frac{r^2}{R^2} + ...)\right\} && (12.58) \\ &= -\frac{2GMm}{r^2}\left\{1 + 3\frac{r^2}{R^2} + ...\right\} && (12.59) \end{aligned}$$

If r is small *enough* relative to R, the force is indeed what we'd expect if all of the mass $2m$ were concentrated at the center of mass at radius R. We see that there is a correction term that scales like r^2/R^2, but for the moment we'll ignore it since $r \ll R$ by assumption.

Since the total force must equal the total mass times the acceleration of the center of mass, we can conclude that the dumbbell will have an acceleration of:

$$F_{\rm tot} = -\frac{2GMm}{R^2} = (2m)a_{\rm cm} \quad \Rightarrow \quad a_{\rm cm} = -\frac{GM}{R^2} \tag{12.60}$$

What does this have to do with the tide? Well, if the *rigid* rod is freely falling, *both masses m must have the same acceleration $a_{\rm cm}$, so both must experience the same downward force!* However, *gravity alone* exerts a *different* force on the two masses. We conclude that there must be some *additional* force exerted on the two masses *by the rod*, pictured as a (red) "tension" T in the figure above.

The directions in this figure make sense. The top mass has a gravitational force down that is *too small* to provide the common acceleration; the bottom one has a gravitational force that is similarly *too large* to provide the common acceleration. The missing force must be made up by *tension in the rod*, acting down on the upper mass, up on the lower one, to transfer some of the surplus force from the lower to make up the deficit of the upper!

Newton's second law for the upper mass is thus (to lowest order, with "up" positive):

$$\begin{aligned} F_{\rm upper} &= -\frac{GMm}{R^2}\left\{1 - \frac{2r}{R} + \cancel{\mathcal{O}\left(\frac{r^2}{R^2}\right)}\right\} - T \quad \Rightarrow \\ ma_{\rm cm} &= -\frac{GMm}{R^2} + \frac{2GMmr}{R^3} - T \\ -\cancel{\frac{GMm}{R^2}} &= -\cancel{\frac{GMm}{R^2}} + \frac{2GMmr}{R^3} - T \end{aligned}$$

or

$$T = \frac{2GMmr}{R^3} \tag{12.61}$$

If you repeat the argument for the lower mass, you will conclude that the tension in the rod there acts *up* with exactly the same magnitude.

If we measure T with (say) a spring attached to the end of the rod, that spring will *stretch*. This stretch is not exactly due to a pseudoforce per se that has appeared in the center of mass reference frame due to its acceleration, but is instead due to the fact that the *real* force of gravity is no longer identical for the upper and lower mass but they are constrained by the rod to move as one. It still behaves a lot *like* a pseudoforce, one that appears to point *away* from the center of mass of the rod at both the upper end of the rod and the lower one, *stretching* the rod. ***This is the tidal force*** in the accelerating center of mass frame, and it scales like $2r/R$ relative to the actual gravitational force acting on the upper mass.

For objects the size of rocket ships outside planets the size of Earth, of course, this correction is one part in around a million or even smaller – surely ignorable and almost impossible to measure. For objects the size of *moons* near *planets* or *planets* near *stars*, however, the tidal force may not be negligible at all!

Things are a bit different for an object (like a moon) in a tidally locked orbit. In that case the acceleration of the center of mass reference frame is *two* dimensional (in a plane) and more complicated. To understand this, we'll next consider the mixture of real forces and pseudo-forces acting on the Moon collectively and apply the results to objects (e.g. blocks of mass m) *on* the Moon in different places as the Moon orbits the Earth in an approximately circular orbit.

Example 12.8.2: The Moon in a Circular Orbit

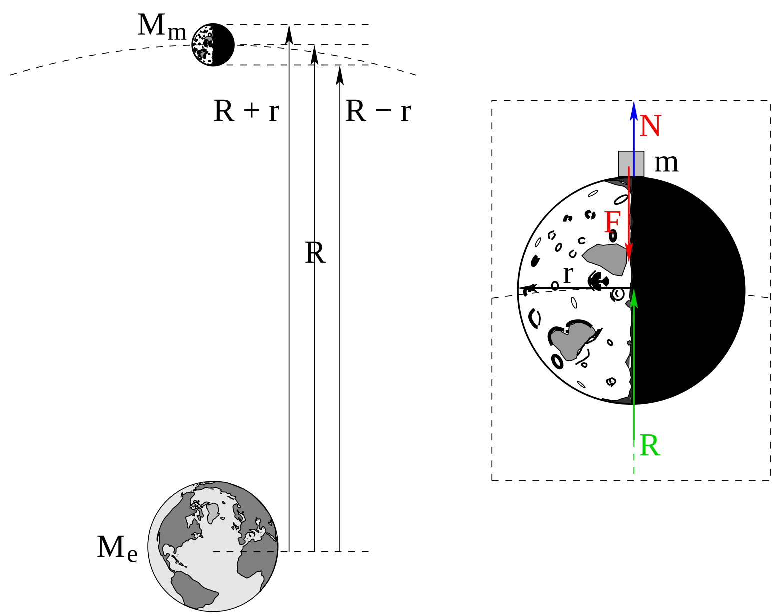


Figure 12.12: The Moon orbiting the Earth in an approximately circular orbit. The diagram is not to scale and the parameters are deliberately chosen to resemble those in the previous example. The inset shows a block of mass m on the surface of the moon on the far side away from the Earth (at a distance $R + r/2$ from the center of the Earth where R is the distance between the centers and r is the radius of the moon).

From the previous example[298] we know that to leading order in r/R, the gravitational force on the moon itself in Newton's Second Law is (with "up" in the figure once again positive) ought to scale like:

$$F_{\rm moon} = -\frac{GM_eM_m}{R^2}\left\{1 + \cancel{\mathcal{O}\left(\frac{r^2}{R^2}\right)}\right\} = M_m a_c = -M_m\Omega_m^2 R \qquad (12.62)$$

[298] Imagine the Moon to consist of a large collection of "dumbbells", for example.

for a circular orbit with angular velocity Ω_m. We'll use this to write:

$$\Omega_m^2 = \frac{GM_e}{R^3}$$

as usual for circular orbits[299].

The moon is *tidally locked* – it always turns the same face towards the Earth. Consider a small mass m placed on the far side of the Moon (at a distance $R + r$ from the center of the Earth and a distance r from the center of the Moon). This mass *also* travels in a circle around the Earth of radius $R + r$! From the inset in figure 12.12, Newton's Second Law for this mass is thus (with "up" in the figure positive):

$$N - \frac{GM_m m}{r^2} - \frac{GM_e m}{(R+r)^2} = -m\Omega_m^2(R+r) \tag{12.63}$$

Recall that the normal force acting on the block is the "apparent weight" of the block in the accelerating frame:

$$N = \frac{GM_m m}{r^2} + \frac{GM_e m}{(R+r)^2} - m\Omega_m^2(R+r) \tag{12.64}$$

As before, let's use the binomial expansion to express:

$$\frac{GM_e m}{(R+r)^2} = \frac{GM_e m}{R^2}\left(1+\frac{r}{R}\right)^{-2} \approx \frac{GM_e m}{R^2}\left(1 - 2\frac{r}{R} + ...\right)$$

Next we substitute the expression we obtained above for Ω_m^2 and simplify:

$$\begin{aligned} N &= \frac{GM_m m}{r^2} + \frac{GM_e m}{R^2} - 2\frac{GM_e mr}{R^3} - \frac{GM_e m}{R^3}(R+r) && (12.65)\\ &= \frac{GM_m m}{r^2} + \cancel{\frac{GM_e m}{R^2}} - 2\frac{GM_e mr}{R^3} - \cancel{\frac{GM_e m}{R^2}} - \frac{GM_e mr}{R^3} && (12.66)\\ &= \frac{GM_m m}{r^2} - 3\frac{GM_e mr}{R^3} = \frac{GM_m m}{r^2} - 3m\Omega_m^2 r && (12.67) \end{aligned}$$

That is (to conclude):

> **Objects on the side of the Moon opposite to the Earth "weigh" less than they would due to the Moon's gravitation alone by $3m\Omega_m^2 r$.**

Furthermore, our derivation of this remarkable result shows the *dual origin* of this force. One part out of three is a pure pseudoforce resulting from the fact that the centripetal acceleration of the mass m must be larger by $\Omega_m^2 r$ than that of the center of mass of the Moon. Two parts are due to the fact that the real force component due to the Earth's gravitational field there is *weaker* than it is at the center (this is the only term that contributed in the previous example).

[299]The effect of the variation of the Earth's gravitational field *does* very slightly increase the force on the moon relative to what we expect from assuming that all of the mass of the moon is concentrated at its center, but:

$$\frac{r^2}{R^2} \approx \frac{(1.7\times10^3)^2}{(3.8\times10^5)^2} \approx 0.00002$$

so the worst case correction (one that assumes that the Moon is a barbell shape instead of a sphere) is of the order of roughly 2 *thousandths of a percent* and is surely much less for a sphere. We will neglect it.

No wonder the tide is complicated! It is a mix of a kinematic term arising from changing to the accelerating frame of the Moon itself and the fact that the Earth's gravitational field is *smaller* instead of *larger* there! A mass m in the center of the Moon *at* the radius R would have the pseudoforce and the actual force precisely cancel, would experience no gravitational force due to the Moon itself, and would be "weightless" in the accelerating frame, in pure free-fall around the Earth in a circular orbit. A mass m on the surface of the Moon at a position a distance R from the center of the Earth would experience its normal "Moon weight" of $GM_m m/r^2 = mg' \approx mg/6$. But what about a mass located at $R - r$, on the face of the Moon *closest* to the Earth?

Obviously with *up still positive*:

$$-N + \frac{GM_m m}{r^2} - \frac{GM_e m}{(R-r)^2} = -m\Omega_m^2(R - r) \tag{12.68}$$

(In other words, N now points "down" towards the Earth and Moon gravity acting on m points "up", but *the force due to the Earth and its centripetal acceleration still point down!* This only alters the binomial expansion

$$\frac{GM_e m}{(R-r)^2} = \frac{GM_e m}{R^2}\left(1 - \frac{r}{R}\right)^{-2} \approx \frac{GM_e m}{R^2}\left(1 + 2\frac{r}{R} + ...\right)$$

and the signs. Now

$$N = \frac{GM_m m}{r^2} - \frac{GM_e m}{R^2} - 2\frac{GM_e mr}{R^3} + \frac{GM_e m}{R^3}(R - r) \tag{12.69}$$

$$= \frac{GM_m m}{r^2} - \cancel{\frac{GM_e m}{R^2}} - 2\frac{GM_e mr}{R^3} + \cancel{\frac{GM_e m}{R^2}} - \frac{GM_e mr}{R^3} \tag{12.70}$$

$$= \frac{GM_m m}{r^2} - 3\frac{GM_e mr}{R^3} = \frac{GM_m m}{r^2} - 3m\Omega_m^2 r \tag{12.71}$$

precisely as before! Objects on the *near* side of the Moon *also* have an apparent weight that is less than expected due to the Moon's gravitation alone, and (to leading order only!) the magnitude of the peculiar mixture of real and pseudoforce giving rise to this is the *same* as it is on the *far* side!

This can be intuitively understood by realizing that the centripetal acceleration towards the Earth relative to that of the center of mass of the Moon is *reduced* by $\Omega_m^2 r$, while the gravitational *field* of the Earth there is *increased* by $2\Omega_m^2 r$. This means N doesn't have to oppose the full weight of the mass m due to the Moon's own gravitation as the Earth is pulling it more and it is in fact accelerating less.

Let's revise our previous conclusion accordingly:

> **Objects on *both* the side of the Moon closest to the Earth *and* the side of the Moon farthest from the Earth weigh less than they would due to the Moon's gravitation alone by the "tidal force" $3m\Omega_m^2 r$. Only objects located on the surface of the Moon on the circle that happens to be "at" the constant distance R from the Earth have a measured weight equal to their true lunar weight.**

Just how *big* is this correction? Well, we know that:

$$\Omega_m = \frac{2\pi}{27.3 \times 86400} = 2.66 \times 10^{-6} \text{ rad/sec}$$

and $r = 1.7 \times 10^6$ meters, hence:

$$a_m = \frac{F_{\rm tide}}{m} = 3\Omega_m^2 r = 3.6 \times 10^{-5}\ {\rm N/kg}$$

or 36 *micro*Newtons per kilogram. Not very big, to be sure, but still $3r/R = 1.3 \times 10^{-2}$ times the acceleration of the moon itself:

$$a_{\rm moon} = \Omega_m^2 R = 2.7 \times 10^{-3}\ {\rm N/kg}$$

Hence $F_m = ma_m$ at either the closest or farthest (from the Earth) points is around a 1% correction to the force that would keep the same mass m falling in a circular orbit at around the Earth at the radius R! Small, but hardly negligible.

The cases we've looked at above are relatively simple (although you might not think so, understandably) because we took great care to line things up and make the problems torque free and one dimensional along r. However, there is one more surprise in store for us. It turns out that when a mass distribution is *not* perfectly aligned with r, the variation in the gravitational field can exert a *nonzero torque*, one that affects both the spin angular momentum of the object (if any) and its orbital angular momentum.

This seems like it might be pretty important, but at the same time, the analysis here is a lot more difficult than the simple examples above because we have to work out triangles using the Law of Cosines and Law of Sines, which you might or might not remember.

For that reason, the example below is considered ***advanced and optional***, not necessarily part of the course. Physics majors or students hoping to work in aerospace engineering or climate modeling or astrophysics should probably work through it in some detail; not so much anybody else unless, of course, you want to skim it to get to the conclusion or use it to practice the underlying math (to make the easier math above *much* easier for you in the long run)

Example 12.8.3: A Tipped, Freely Falling Dumbbell

Suppose we have a massless rod of length $2r$ in free fall with two identical masses m at the ends, tipped at an arbitrary angle θ relative to the vector $\vec{R}$ from (say) the Earth's center of attraction to the center of mass of the rod, in a case where the angle θ and r is *not necessarily* particularly small (at first) relative to R. This is pictured in figure 12.13.

Even though this is just our first example above tipped to an angle θ, things are now *a lot more complicated!* We can easily find the vector coordinates in the provided xy coordinate frame for $\vec{r}_1$. It is a lot more work to find the components of $\vec{F}_1$, and it is a bit difficult to make out all of the geometry in the first figure[300].

[300] From here on, everything we do with the left hand mass 1 can be done identically for the right hand mass 2 by changing a few signs, so we will just solve for the force and torque on mass 1 and then *write down the equivalent result* for mass 2.

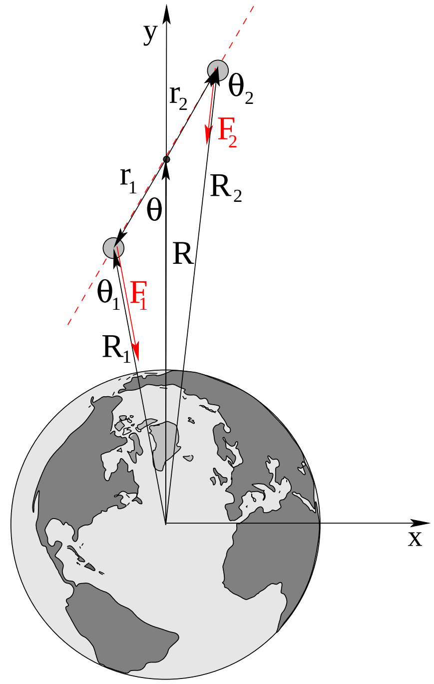
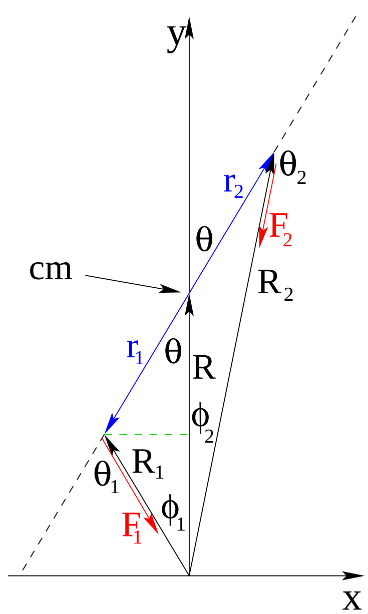


Figure 12.13: A massless rod of length $2r$ with identical point masses m located at the ends is located a distance $r < R$ away from a planetary center, turned to a small angle relative to the direction of $\vec{\boldsymbol{R}}$. Note that the masses themselves are at vector positions $\vec{\boldsymbol{r}}_1$ and $\vec{\boldsymbol{r}}_2$ (both with magnitude r) relative to the center of mass of the rod-mass system, and at vector positions $\vec{\boldsymbol{R}}_1$ and $\vec{\boldsymbol{R}}_2$ respectively relative to the center of the planet.



To simplify our job, then, I present an "exagerrated" picture of the triangles and angles involved to the left. We will have to use a substantial amount of trigonometry, in particular the Law of Cosines and Law of Sines, in order to find F_{1x} and F_{2x}, *especially because we want to get those components in terms of* m*,* M*,* R*,* r*, and* θ *only!*

For the left hand mass, we need to find ϕ_1 and R_1 in order to be able to find the total vector torque and force in terms of these "givens". We'll start with the Law of Cosines, noting again that $|\vec{\boldsymbol{r}}_1| = |\vec{\boldsymbol{r}}_2| = r$:

$$R_1 = \sqrt{R^2 + r^2 - 2Rr\cos\theta} \tag{12.72}$$

We can then use the Law of Sines to connect:

$$\frac{\sin\phi_1}{r} = \frac{\sin\theta}{R_1} \quad \Rightarrow \quad \sin\phi_1 = \frac{r\sin\theta}{R_1} = \frac{r\sin\theta}{\sqrt{R^2 + r^2 - 2Rr\cos\theta}} \tag{12.73}$$

which we will need to find F_{1y}. To find F_{1x} we will similarly need $\cos\phi_1$:

$$\begin{aligned}
\cos\phi_1 &= \sqrt{1 - \sin^2\phi_1} && (12.74)\\
&= \sqrt{1 - \frac{r^2\sin^2\theta}{R^2 + r^2 - 2Rr\cos\theta}} && (12.75)\\
&= \sqrt{\frac{R^2 + r^2 - 2Rr\cos\theta \ - \ r^2\sin^2\theta}{R^2 + r^2 - 2Rr\cos\theta}} && (12.76)\\
&= \sqrt{\frac{R^2 + r^2\cos^2\theta - 2Rr\cos\theta}{R^2 + r^2 - 2Rr\cos\theta}} && (12.77)
\end{aligned}$$

These three results in terms of the givens are perhaps not "simple", but they are easy enough to compute with.

With them in hand, evaluating the components of $\vec{r}_1$ and $\vec{F}_1$ is now straightforward. We get:

$$\begin{aligned}
r_{1x} &= r\sin\theta && (12.78)\\
r_{1y} &= -r\cos\theta && (12.79)
\end{aligned}$$

and

$$\begin{aligned}
F_{1x} &= \frac{GMm}{R_1^2}\sin\phi_1 = \frac{GMm\,r\sin\theta}{(R^2 + r^2 - 2Rr\cos\theta)^{3/2}} && (12.80)\\
F_{1y} &= -\frac{GMm}{R_1^2}\cos\phi_1 = -\frac{GMm\,(R^2 + r^2\cos^2\theta - 2Rr\cos\theta)^{1/2}}{(R^2 + r^2 - 2Rr\cos\theta)^{3/2}} && (12.81)
\end{aligned}$$

A quick check: We should get the same result we got in the first example for the lower mass if we set $\theta = 0$ so $\sin\theta = 0$ and $\cos\theta = 1$:

$$\begin{aligned}
F_{1x} &= 0 && (12.82)\\
F_{1y} &= -\frac{GMm\,(R^2 + r^2 - 2Rr)^{1/2}}{(R^2 + r^2 - 2Rr)^{3/2}} = -\frac{GMm}{(R-r)^2} && (12.83)
\end{aligned}$$

exactly as expected!

It is hopefully obvious how to find the similar components of $\vec{F}_2$ and $\vec{r}_2$ for the upper mass – basically it involves changing one sign to get R_2:

$$R_2 = \sqrt{R^2 + r^2 + 2Rr\cos\theta} \tag{12.84}$$

as well as the signs of the components in terms of ϕ_2 on that side. I'm leaving this, and the explicit summing up of the *total* force and torque acting on the system, to the interested student as an exercise in algebra.

The last few things to note in this example can be read right off of the original figure 12.13. Just by inspection, it is obvious that the resultant force vector is no longer parallel to $\vec{R}$! In *any* frame, and whether or not the center of mass of the object is initially released from rest or is in an orbit of some sort, the position-dependent gravitional field is exerting a ***nonzero torque*** on the object so that *the angular momentum of the subsystem consisting of the two masses and rod is not contant in time!*

With a great deal of work doing binomial expansions, one can still show that in the $r \ll R$ limit:

$$F_{\rm tot} = -\frac{2GMm}{R^2} = 2ma$$

"down", as $\phi_{1,2} \to 0$ as $r/R \to 0$. The rod will still carry a non-zero tension, and even though the rod is symmetric in mass, *center of gravity* is no longer located at the *center of mass* for the system – it is closer to the central attractor than the center of mass. The free motion is going to be very complicated indeed, as it will *want* to oscillate in θ and orbit at the same time, with some coupling between the two.

With this complex a result for what is after all, arguably the simplest *three* dimensional system we could think up with a easily located center of mass and a simple, symmetric geometry, you can imagine how difficult it would be to compute the total force or torque acting on a potato-shaped asteroid orbiting another asteroid or (for that matter) the total force acting on the Moon as it orbits the Earth (even ignoring the Sun around which both are *also* orbiting)[301].

12.8.1: Earth Tides

Our final chore in this section is to discuss the tides on *Earth*. The Earth experiences significant tidal forces from *both* the Moon *and* the Sun, with lunar tides stronger and dominant over solar tides. From our analysis above, we expect that things will weigh *less than they should due to Earth gravity alone* when they are *closest* to the Moon during the day or *farthest* from the Moon during the day, with an extra modulation that enhances this effect when the Sun, Moon and Earth form a line, (new moon or full moon) and minimizes this effect when the Sun is at right angles to the Moon in the orbit of the latter – first or last quarter. This is graphically represented in figure 12.14[302].

The maximum oceanic tides produced when the Sun, the Earth, and the Moon are aligned are called ***spring tides***, and occur approximately twice a month at full Moon and new Moon. The minimum tides in the first and last quarter (when half of the Moon is dark as seen from the Earth) are called ***neap tides*** and also occur approximately twice a month. Every day there are approximately two high tides and two low tides, both separated by around twelve and a half hours (so that the time of the high tides advances by a bit less than an hour a day). Coastal high tides do *not* generally occur at any given spot when the moon is directly overhead – there usually a significant phase shift.

[301]Can you say "***Rocket Scientist***"? This is why engineers and physicists and programmers working for NASA make the big bucks. It actually isn't at all impossible to write *computer codes* to do all of the bookkeeping and solve the general physics problem, but it sure as heck isn't simple enough to even make a decent *homework problem* in an intro course like this one!

[302]But don't take this figure too seriously – the average tidal bulge of the ocean is order of a meter over most of the Earth, although near land many factors such as wind, the shape of the land, and latitude can cause it to over ten meters – see e.g. the Wikipedia: http://www.wikipedia.org/wiki/Bay of Fundy, where the highest tides on Earth occur and one can actually ride rivers *backwards* as the tidal flood pushes deeply inland!

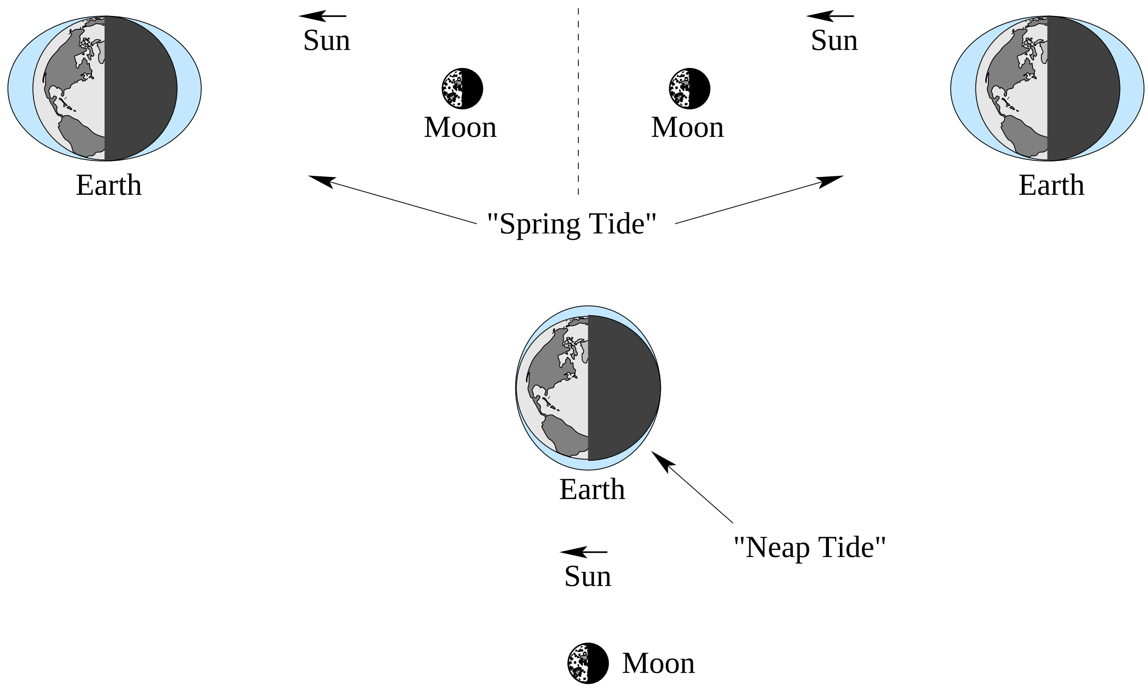


Figure 12.14: The largest ("Spring") tides occur roughly twice a month when the Sun, the Earth, and the Moon are in an approximately straight line – new moon and full moon. The smallest ("Neap") tides occur roughly twice a month when the Sun, the Earth, and the Moon form a right angle – first quarter and last quarter. Note that the ***Moon*** dominates the tidal bulge – the Sun alone would make much smaller tides!

This lowering of the *effective* weight applies to *everything* – the air, the water, and the soil and rock of the Earth itself! Since all of this matter is at least *somewhat* plastic, it *all* moves up and down roughly twice a day as the Earth turns beneath the more slowly advancing moon, although the solunar gravitational tide may not be the dominant dynamical factor

For example, the atmosphere *has* tides, but it turns out that the "tidal" bulge is dominated by its diurnal *heating and cooling* and not solunar gravitational tides, and hence has a periodicity and phase that is different from that expected on the basis of the Sun's and the Moon's locations.

The Earth itself, on the other hand, is subject to the tidal bulge all the way down to its center. Although it appears to be "rock solid" at the crustal surface, it is actually quite pliant across its entire depth! As a consequence, the *Earth itself* lifts and falls by order of a meter twice a day, dominated as expected by the Moon but with a significant contribution from the Sun as well[303]. A nice review of this is in Wikipedia: http://www.wikipedia.org/wiki/Earth_tides if you wish to read further.

The Earth tidal bulge is almost impossible to "notice" with human senses or even detect with ordinary instrumentation as the rise and fall is very gradual and its meter or so of range is spread out over thousands of kilometers. However, it does have some *profound effects* that

[303]Your mileage may vary! Just how far the tides move the surface – and the *direction* they move it, as there can be a sideways component, not just vertical – depends on the position of Sun and Moon, just where you are on the Earth, what the *crust* looks like there (mountain or sea bottom) and can be significantly affected by e.g. the solunar *water* tides at coastal areas where high tides are generally not in phase with lunar tides!

are worth at least looking at and that are indeed experimentally verified.

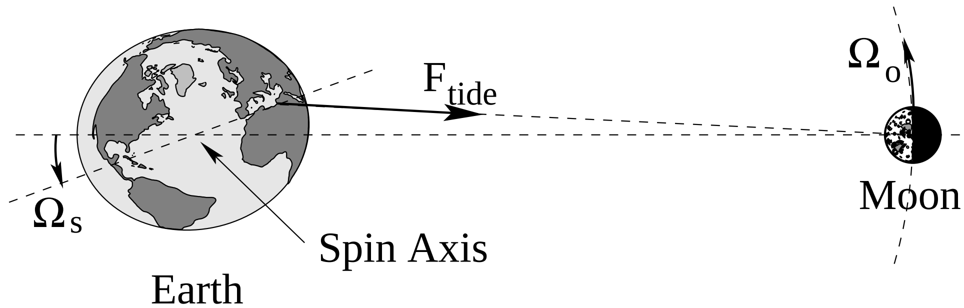


Figure 12.15: The Earth's tidal bulge *leads* the position of the moon because the Earth rotates (spins) on its axis faster than the Moon revolves (orbits) around the Earth on a common axis. This causes the Moon to exert a more or less *constant torque* on the Earth!

One of these is (incompletely) illustrated in figure 12.15. In this figure, the Earth spins on its axis with an angular velocity Ω_s (roughly) out of the page as drawn (so that Moon and Sun are seen to "rise" in the *east* as the Earth rotates). The Moon orbits the Earth with a *smaller* angular velocity $\Omega_o \sim \Omega_s/27$. This means that the Earth Tide bulge facing the moon is constantly being carried *past* the line between the center of the Earth and the Moon. It takes *time* to relax back to the unbulged position, so that the bulge axis is rotated spinward (east) relative to the Moon's position at pretty much all times.

As a consequence, the force of gravitation due to the Moon is no longer directed along the line connecting the centers of Earth and Moon. This force exerts a (comparatively) *small but nonzero torque* on the Earth *into the page*, one that *slows* the Earth's angular spin *out of* the page. As a consequence, Earth days are getting *just a tiny bit longer*[304] all of the time because of the mix of lunar and solar torque (both tend to slow its rotation). This effect is small – so small that it takes roughly 50,000 *years* for the day to become one whole second longer – but not so small that it cannot be measured, as it is order of twenty nanoseconds per year and hence *precisely* the right size to screw up GPS coordinate computations by meters per year if not corrected for!

What about the Moon? If the torque is reducing the Earth's angular momentum, it must be *increasing* the Moon's angular momentum! This has the effect of slightly increasing the radius of the Moon's orbit. In turn, this *also* slightly increases the total mechanical energy of the Moon's orbit (because $E_{\rm tot} = U/2$ for circular orbits and U becomes slightly less negative)! One of several ways to understand this is that some of the Earth's rotational kinetic energy is transferred to the moon in a very, very slow partly inelastic "collision" mediated by the tidal forces. Note well that some of the Earth's initial rotational kinetic energy is also converted into *heat* as the tidal bulge constantly inelastically deforms the rock of the Earth itself. The total mechanical energy of the system (very) slowly decreases.

One of the scientific goals of the very first Apollo missions was to enable precise ranging of the Earth-Moon distance by placing laser reflectors at the lunar landing sites. This distance has been more or less continuously sampled over the last fifty-odd years, and it is now well-established that the Moon does indeed increase the mean radius of its orbit around the Earth by around 3.8 centimeters a year.

[304]Roughly ***2 milliseconds per century***, in fact

This isn't much – recall that the Moon's orbit is approximately 3.84×10^8 meters in radius, so it is one part in 10 billion, invisible to human senses on anything less than geologic time where humans live at most 100 years or thereabouts, barely long enough for the orbit to increase by 1 part in 100 million! However, 500 million years ago, when animal life in the sea had appeared and evolution was exploding, the Moon would have been roughly 5% closer to the Earth[305] than it is now, *visibly* larger in the sky, and total solar eclipses were likely to have been a regular occurence, instead of rare as they are now! The tides themselves would have been larger as well.

Right now the Moon, seen from the Earth, subtends almost exactly the same angle as the Sun does. Consequently total lunar eclipses are possible where the Moon completely covers the Sun's face, although only for a matter of minutes. At the same time, even now many eclipses are annular, where the sun is visible all the way around the moon even at maximum totality. Over the next few thousand years, we can predict that total solar eclipses where the Sun's face is completely covered by the Moon *somewhere* on Earth will grow less and less common, and in the not terribly distant future – in geological time – disappear altogether. Eclipses will still occur, but they will never again completely cover the Sun's face.

We live in a remarkable time. Total solar eclipses still *barely* occur with some regularity, enabling quite a few of the key measurements and observations that ushered in the modern scientific era. The Earth *happens* to have a "North Star" – Polaris – at the end of the "handle" of the little dipper, but the torque exerted by the Sun, the Moon, and the other planets are inexorably causing the axis of rotation to precess, so that in as little as a few *hundred* years Polaris will no longer lie almost exactly on the axis of rotation. The actual gravitational field of the Earth is being mapped to very high precision over most of its face by means of satellites, and we have observed a definite correlation between tidal deformation and seismic activity.

One by one the secrets of what our remote ancestors saw when they looked up at the sky at all of the itty bitty lights that seemed to just hang there, suspended, have yielded to the steady and industrious application of the methods for acquiring epistemologically sound knowledge pioneered during the Enlightenment and launched by the work we have studied due to Galileo, Toricelli, Newton, Cavendish, and many more. Thus far, however, we have almost completely ignored the *other* force – besides that of gravitation – that dominates our quotidian lives: Electromagnetism! Perhaps it is time to give *it* a look!

12.9: Bridging the Gap: Coulomb's Law and Electrostatics

This concludes our treatment of basic mechanics. Gravitation is our first actual *law of nature* – a force or energy law that describes the way we think the Universe actually works at a

[305]Estimate: $5 \times 10^8 \times 0.038 = 1.9 \times 10^7$ meters or 19,000 kilometers, or 5% of the current radius of 3.8×0^5 kilometers. The tides would have been even larger then as well!

fundamental level.

Gravity is, as we have seen, important in the sense that we live gravitationally bound to the outer surface of a planet that is itself gravitationally bound to a star that is gravitationally compressed at its core to the extent that thermonuclear fusion keeps the entire star white hot over billions of years, providing us with our primary source of usable energy. It is *unimportant* in the sense that it is *very weak*, the weakest of all of the known forces.

Next, in the second volume of this book, you will study one of the *strongest* of the forces, the one that dominates almost every aspect of your daily life. It is the force that binds atoms and molecules together, mediates chemistry, permits the exchange of energy we call light, and indeed is the fundamental source of *nearly every* of the "forces" we treated in this semester in collective form: The electromagnetic interaction.

Just to whet your interest (and explain why we have spent so long on gravity when it is weak and mostly irrelevant outside of its near-Earth form in everyday affairs) let is take note of ***Coulomb's Law***, the force that governs the all-important electrostatic interaction that binds electrons to atomic nuclei to make atoms, and binds atoms together to make molecules. It is the force that exists between two *charges*, and can be written as:

$$\vec{\boldsymbol{F}}_{12} = \frac{k_e \ q_1 \ q_2}{r_{12}^2} \hat{\boldsymbol{r}}_{12} \tag{12.85}$$

Hmmm, this equation looks rather familiar! It is *almost identical* to Newton's Law of Gravitation, only it seems to involve the *charge* (q) of the particles involved, not their mass, and an *electrostatic constant* k_e instead of the gravitational constant G.

In fact, it is *so* similar that you instantly "know" lots of things about electrostatics from this one equation, plus your knowledge of gravitation. You will, for example, learn about the electrostatic field, the electrostatic potential energy and potential, you will analyze circular orbits, you will analyze trajectories of charged particles in uniform fields – all pretty much the same idea (and algebra, and calculus) as their gravitational counterparts.

The one really *interesting* thing you will learn in the first couple of weeks is how to *properly* describe the geometry of $1/r^2$ force laws and their underlying fields – a result called ***Gauss's Law***. This law and the other Maxwell Equations will turn out to govern nearly everything you experience. In some very fundamental sense, you *are* electromagnetism.

Good luck!

Homework for Week 12

Problem 1.

Physics Concepts: Make this week's physics concepts summary as you work all of the problems in this week's assignment. Be sure to cross-reference each concept in the summary to the problem(s) they were key to, and include concepts from previous weeks as necessary. Do the work carefully enough that you can (after it has been handed in and graded) punch it and add it to a three ring binder for review and study come finals!

Problem 2.

It is a misconception that astronauts in orbit around the Earth are ***weightless***. *Weight* (recall) is a measure of the *actual gravitational force exerted on an object*, something that is *not* zero when you are in orbit!

Suppose you are in a space shuttle orbiting the Earth at a distance of two times the Earth's radius ($R_e = 6.4 \times 10^6$ meters) from its center.

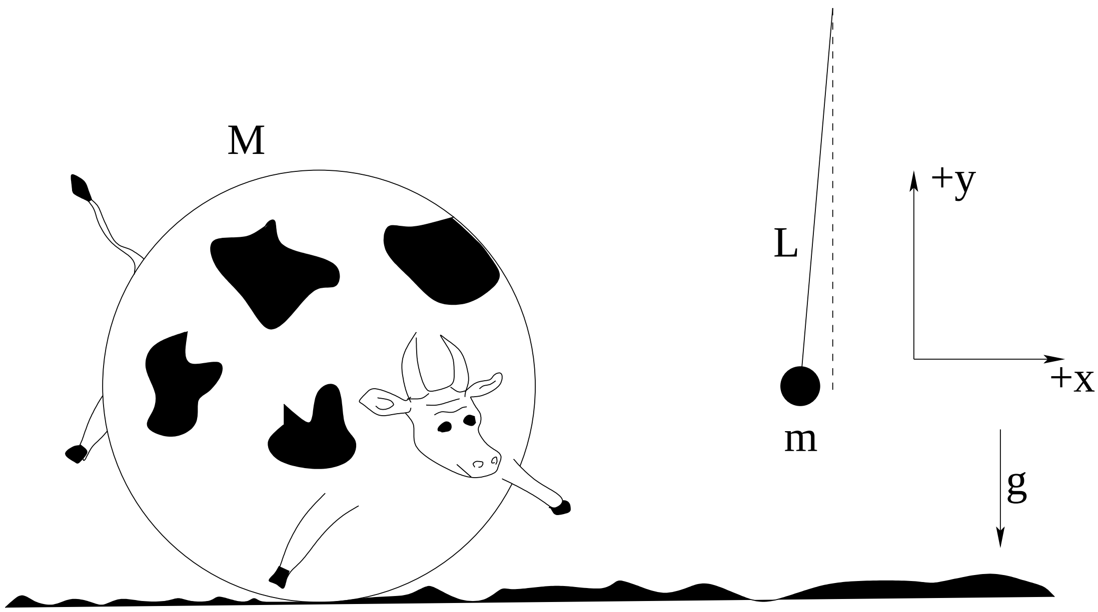
a) What is your weight *relative* to your weight on the Earth's surface?

b) Does your weight depend on whether or not you are moving at a constant speed? Does it depend on whether or not you are accelerating?

c) Why would you *feel* weightless inside an orbiting shuttle?

d) Can you ***feel*** as "weightless" as an astronaut on the space shuttle (however briefly) in your own dorm room? How?

Problem 3.



There is an old physics joke involving cows[306] , and you will need to use its punchline to solve this problem.

A cow is standing in the middle of an open, flat field. A plumb bob with a mass of 1 kg is suspended via an unstretchable string 10 meters long so that it is hanging down roughly 2 meters away from the center of mass of the cow. *Making any reasonable assumptions you like* (one is illustrated for you), ***estimate*** the ***angle of deflection of the plumb bob from vertical*** due to the gravitational field of the cow.

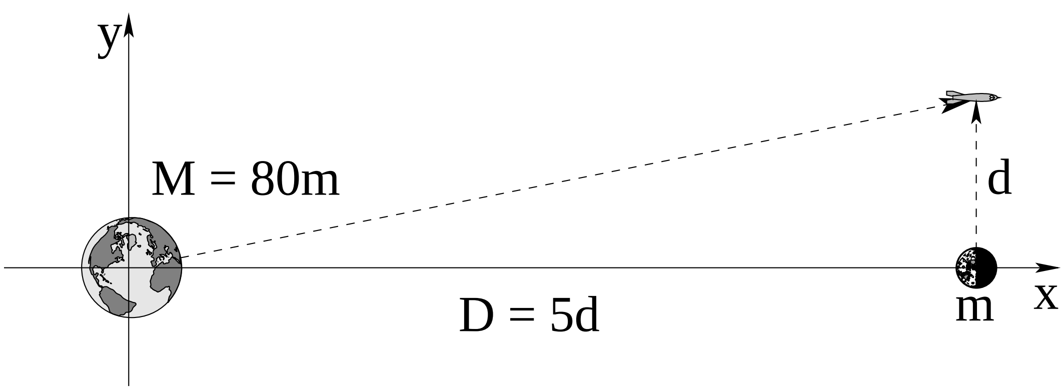
[306]Wikipedia: http://www.wikipedia.org/wiki/Spherical cow. Naturally, it actually *has its own Wikipedia page*, so you can read it and enjoy...

Problem 4.

Physicists are working to understand "dark matter", a phenomenological hypothesis invented to explain the fact that things such as the orbital periods around the centers of *galaxies* cannot be explained on the basis of estimates of Newton's Law of Gravitation using the total *visible* matter in the galaxy (which works well for the mass we can see in planetary or stellar context). By adding mass we cannot see until the orbital rates are explained, Newton's Law of Gravitation is preserved (and so are its general relativistic equivalents).

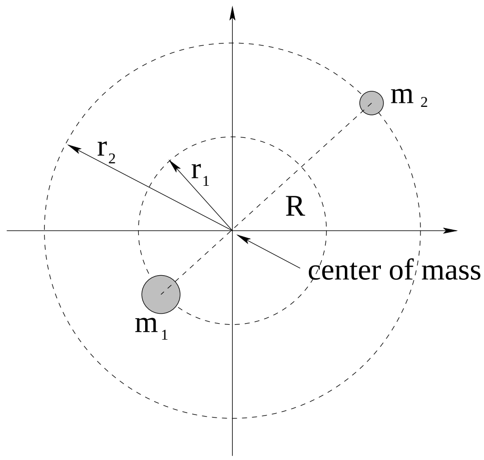
However, there are alternative hypotheses, one of which is that Newton's Law of Gravitation is *wrong*, deviating from a $1/r^2$ force law at very large distances (but remaining a central force). The orbits produced by such a $1/r^n$ force law (with $n \neq 2$) would not be elliptical any more, and $r^3 \neq CT^2$ – but would they still sweep out equal areas in equal times? Explain.

Problem 5.



Find the ***vector gravitational field*** acting on the spaceship on its way from Earth to Mars (swinging past the Moon at the instant drawn) in the picture above. Express it in cartesian coordinates and draw it in on a (copy of) the figure above.

Problem 6.



In the discussion of gravitation and orbits in the text, we have implicitly assumed that one of the two objects – the Sun in the case of planetary orbits or the planet (e.g. Earth) in the case of satellite orbits – has a much greater mass than the other. In this case, the center of mass of the system is "inside" the larger object and we can pretend that it remains at rest while the lighter one orbits it.

In reality, though, both objects are in opposing orbits around the *center of mass* of the two objects. In this problem, you will try to figure out what happens if the two objects have *similar* masses.

Suppose two stars, a lighter one with mass m_2 and a heavier one with mass $m_1 = 2m_2$ are each orbiting their mutual center of mass in ***circular orbits*** with radii r_1 and r_2 respectively as drawn above. Answer the following questions as you analyze their orbits:

a) Suppose $R = r_1 + r_2$ is the distance between the two stars. What are r_1 and r_2 in terms of R?

b) What is the ***magnitude*** of the gravitational force F_1 acting on mass m_1? Is the magnitude of the force F_2 acting on m_2 the same or different?

c) The two stars are far away from all other masses so that there is no net external force on them from objects *outside* of this system. The center of mass (in the center of mass reference frame illustrated above) remains at rest. Using this, is Ω_1, the angular velocity of mass m_1 the same or different from Ω_2, the angular velocity of mass m_2?

d) Using your answers to part b), write ***Newton's Second Law*** for each mass. Express the radial acceleration a_i of each mass in terms of Ω_i, the angular velocity of the mass (which may or may not be the same for both masses) and r_i, the radius of its circular orbit.

e) Add the two equations and show that:

$$T^2 = \frac{4\pi^2}{G(m_1 + m_2)} R^3 \qquad \text{for either/both of the planets.}$$

Problem 7.

Recapitulating the text and class/lecture example, in a few lines prove Kepler's third law:

$$r^3 = CT^2$$

for *circular* orbits around a planet or star of mass M and and determine the constant C. Then use it to answer the following questions:

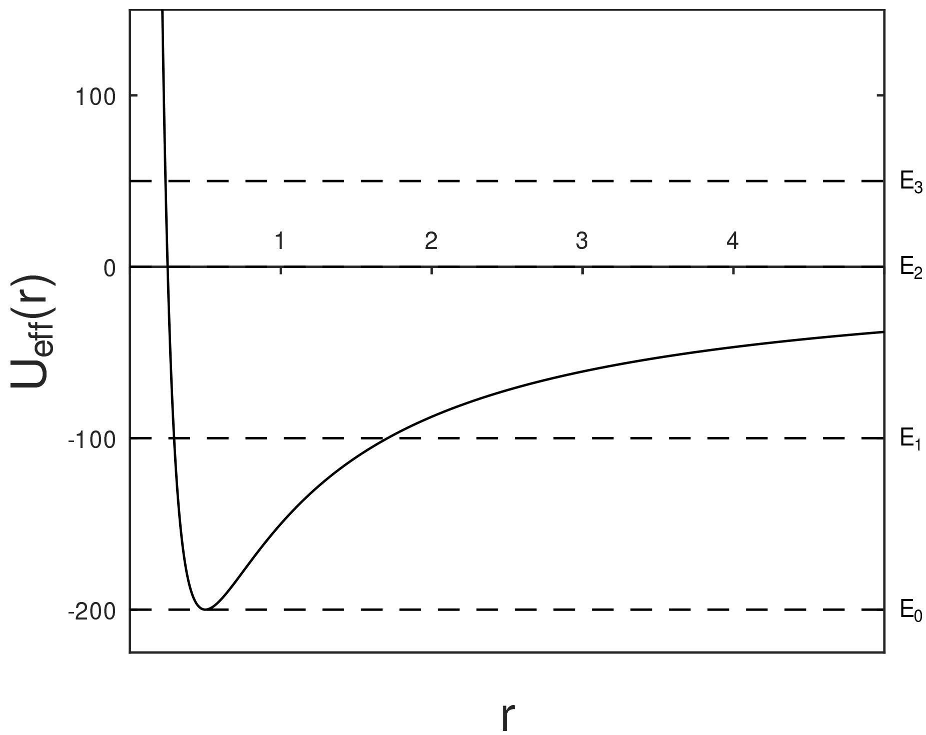
a) Jupiter has a mean radius of orbit around the sun equal to 5.2 times the radius of Earth's orbit. How long does it take Jupiter to go around the sun (what is its orbital period or "year" T_J)?

b) Given the distance to the Moon of 3.84×10^8 meters and its (sidereal) orbital period of 27.3 days, find the mass of the Earth M_e.

c) Using the mass you just evaluated and your knowledge of g on the surface, estimate the radius of the Earth R_e.

Check your answers using google/wikipedia. Think for just one short moment how much of the physics you have learned this semester is verified by the correspondance between theory and published (experimental) results. Remember, I don't want you to believe anything I am teaching you because of my *authority* as a teacher but because it ***works to explain observations***.

Problem 8.

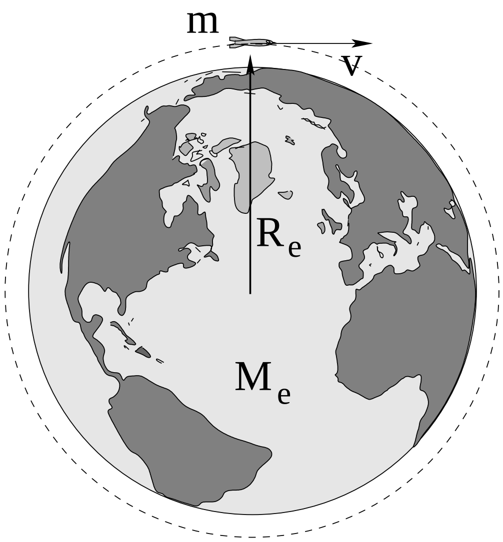


The *effective radial potential* of a planetary object of mass m in an orbit around a star of mass M is:

$$U_{\rm eff}(r) = \frac{L^2}{2mr^2} - \frac{GMm}{r}$$

The total energy of four orbits are drawn as dashed lines on the figure above displaying $U_{\rm eff}$ for some given values of the parameters. Name the kind of orbit (circular, elliptical, parabolic, hyperbolic) for each energy E_i on the figure above and on (a copy of) the figure, draw vertical lines to mark and find each orbit's turning point(s).

Problem 9.



It is very costly (in energy) to lift a payload from the surface of the earth into a circular orbit, but once you are there, it only costs you that same amount of energy again to get from that circular orbit to anywhere you like – if you are willing to wait a long time to get there.

Science Fiction author Robert A. Heinlein succinctly stated this as: "**By the time you are in orbit, you're halfway to anywhere.**"

Prove this by comparing the **total mechanical energy** of a rocket of mass m:

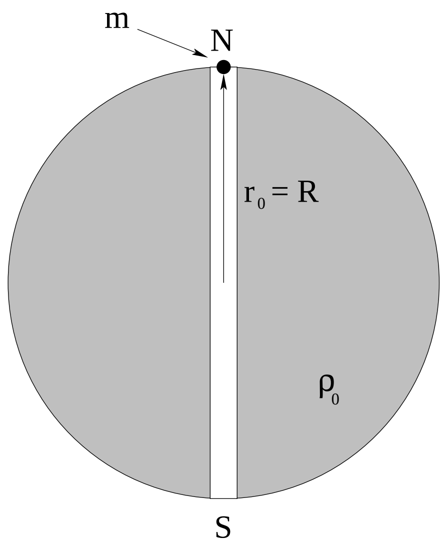
a) On the ground. Neglect its kinetic energy due to the rotation of the Earth although it isn't *quite* negligible.

b) In a (very low) circular orbit with at radius $r \approx R_e$ – assume that it is still more or less the same distance from the center of the Earth as it was when it was on the ground.

c) The orbit with minimal escape energy (that will arrive, at rest, "at infinity" after an infinite amount of time). Show how these three results "prove" Heinlein's assertion to be true in the sense given above.

d) While you are analyzing circular orbits *find the period* T_e *of low-Earth orbit* when $r \approx R_e$ in terms of the mass M_e of the earth, R_e, and any other needed constants. You will need this for another homework problem and the practice is useful!

Problem 10.



A straight, smooth (frictionless) transit tunnel is dug through a planet of radius R whose mass density ρ_0 is constant. The tunnel passes through the center of the planet and is lined up with its axis of rotation (so that the planet's rotation is **irrelevant** to this problem). All the air is evacuated from the tunnel to eliminate drag forces.

a) Find the force acting on a car of mass m a distance $r < R$ from the center of the planet.

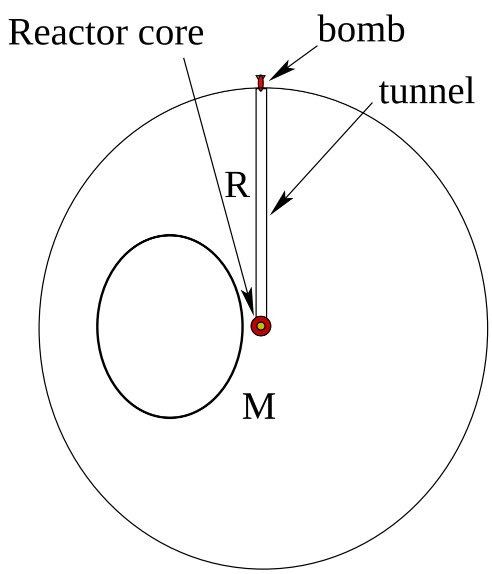
b) Write Newton's second law for the car, and extract the differential equation of motion. From this find $r(t)$ for the car, assuming that it ***starts at rest*** at $r_0 = R$ on the North Pole at time $t = 0$.

c) How long does it take the car to go all the way to the South Pole and then back to the North Pole? Compare this time to the ***period of a circular orbit*** around a star "just over" the surface of the planet that you found (disguised as Ω) in the previous problem.

d) Suppose the mass is released at rest from an initial position $r_0 = R/2$ (halfway to the center) instead of from r_0. How long does it take for the mass to get to the center? Does the answer depend on r_0 (as long as $r_0 \leq R$)?

All answers should be given in terms of G, ρ_0, R and m.

Problem 11.



Death Star Plans

A flaw has been found! A straight, smooth (frictionless) transit tunnel is in the plans of a spherical asteroid of radius R and mass M that has been converted into Darth Vader's ***Death Star***. The tunnel begins at the "north pole" and goes all the way to the *reactor core* at the center of the death star! The death star moves about in a hard vacuum, of course, and the tunnel is open so there are no drag forces or friction acting on masses moving straight down it!

A squad has managed to land on the surface of the Death Star just outside the entrance to the tunnel. They are about to drop an unpowered thermonuclear bomb down the tunnel so that it can set up a reaction in the core and distroy this menace to all that is good and right and make a galaxy far far away safe for the masses of sentient beings that inhabit it. The squad would *like* to have time to get away from the Death Star once they release the bomb, but are not sure how long it will take to reach the reactor core.

a) Help them out! Find an expression for the ***time required for the bomb to freely fall the distance R from the surface to the center***. You may assume that the Death Star still has a nearly uniformly distributed spherical mass M and that the Death Star itself is not accelerating or rotating.

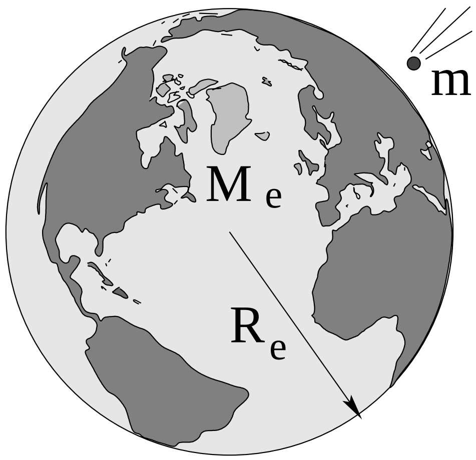
b) The technical plans *do* contain the following data:

$$R = 10^6 \text{ meters} \qquad M = 1.5 \times 10^{22} \text{ kilograms}$$

Evaluate your answer and express it in minutes. Is dropping the bomb down the tunnel like this a good plan?

c) Do they really even *need* a nuclear bomb (which is surely easier to detect than a passive chunk of rock or metal)? Suppose they just drop a deactivated droid with mass $m = 1000$ kg down the tunnel. What would its ***kinetic energy be*** as it reaches the reactor core and smashes into it (*inelastically!*) after falling 1000 km?

Problem 12.



A bitter day comes: a roughly ***spherical*** asteroid of radius R_a and density ρ is discovered that is falling in from ***far away*** so that it will strike the Earth. Ignore the gravity of the Sun in this problem. Your job is to figure out what will happen!

a) First, let's "build an asteroid". Find an algebraic expression for m of the asteroid and evaluate it numerically given that:

$$R_a = 10^4 \text{ meters} \qquad \rho = 6 \times 10^3 \text{ kg/m}^3$$

From now on, you can use m in your algebra and only substitute this number is later as needed.

b) What is the *minimum* possible kinetic energy as it hits the surface of the Earth? Express your ***algebraic*** answer in terms of R_e, either g or G, M_e, and m as you prefer or needed.

c) If it strikes the Earth roughly what fraction of its kinetic energy at the time of impact will be liberated as "heat" or "blast"?

d) Evaluate your answer for the blast energy for the following data:

$$R_e = 6.4 \times 10^6 \text{ meters} \qquad M_e = 6 \times 10^{24} \text{ kilograms}$$

Presumably you know both g and G, so you may not need both of these numbers, and you should use the value of m you computed in a). An approximate/estimate (like the one done in class) is OK.

Express your last answer ***both*** in joules and in (explosive equivalent of) ***tons of TNT*** where the conversion factor is 1 ton-of-TNT = 4.2×10^9 joules. Compare the answer to (say) 30 Gigatons as a safe upper bound for the total combined explosive power of every weapon (including all the ***nuclear*** weapons) on Earth.