Mathematical Methods for Introductory Physics

by

Robert G. Brown

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

Notice

This is a "lecture note" style textbook, designed to support my personal teaching activities at Duke University, in particular teaching its Physics 41/42, 53/54, or 61/62 series (Introductory Physics for potential physics majors, life science majors, or engineers respectively). It is freely available in its entirety in a downloadable PDF form or to be read online at:

http://www.phy.duke.edu/~rgb/Class/intro_physics_2.php

and will be made available in an inexpensive print version via Lulu press as soon as it is in a sufficiently polished and complete state.

In this way the text can be used by students all over the world, where each student can pay (or not) according to their means. Nevertheless, I am hoping that students who truly find this work useful will purchase a copy through Lulu or Amazon when that option becomes available, if only to help subsidize me while I continue to write inexpensive textbooks in physics or other subjects.

Although I no longer use notes to lecture from (having taught the class for decades now, they are hardly necessary) these are 'real' lecture notes and are organized for ease of presentation and ease of learning. They do not try to say every single thing that can be said about each and every topic covered, and are hierarchically organized in a way that directly supports efficient learning.

As a "live" document, these notes have errors great and small, missing figures (that I usually draw from memory in class and will add to the notes themselves as I have time or energy to draw them in a publishable form), and they cover and omit topics according to *my own* view of what is or isn't important to cover in a one-semester course. Expect them to change without warning as I add content or correct errors. Purchasers of any eventual paper version should be aware of its probable imperfection and be prepared to either live with it or mark up their *own* copies with corrections or additions as need be (in the lecture note spirit) as I do mine. The text has generous margins, is widely spaced, and contains scattered blank pages for students' or instructors' own use to facilitate this.

I cherish good-hearted communication from students or other instructors pointing out errors or suggesting new content (and have in the past done my best to implement many such corrections or suggestions).

Books by Robert G. Brown

Physics Textbooks

- *Introductory Physics I and II*

  A lecture note style textbook series intended to support the teaching of introductory physics, with calculus, at a level suitable for Duke undergraduates.

- *Classical Electrodynamics*

  A lecture note style textbook intended to support the second semester (primarily the dynamical portion, little statics covered) of a two semester course of graduate Classical Electrodynamics.

Computing Books

- *How to Engineer a Beowulf Cluster*

  An online classic for years, this is the print version of the famous free online book on cluster engineering. It too is being actively rewritten and developed, no guarantees, but it is probably still useful in its current incarnation.

Fiction

- *The Book of Lilith*

  ISBN: 978-1-4303-2245-0
  Web: http://www.phy.duke.edu/∼rgb/Lilith/Lilith.php

  Lilith is the *first* person to be given a soul by God, and is given the job of giving all the things in the world souls by loving them, beginning with Adam. Adam is given the job of making up rules and the definitions of sin so that humans may one day live in an ethical society. Unfortunately Adam is weak, jealous, and greedy, and insists on being on *top* during sex to "be closer to God".

  Lilith, however, refuses to be second to Adam or anyone else. *The Book of Lilith* is a funny, sad, satirical, uplifting tale of her spiritual journey through the ancient world soulgiving and judging to find at the end of that journey – herself.

- *The Fall of the Dark Brotherhood*

  ISBN: 978-1-4303-2732-5
  Web: http://www.phy.duke.edu/∼rgb/Gods/Gods.php

  A straight-up science fiction novel about an adventurer, Sam Foster, who is forced to flee from a murder he did not commit across the multiverse. He finds himself on a primitive planet and gradually becomes embroiled in a parallel struggle against the world's pervasive slave culture and the cowled, inhuman agents of an immortal of the multiverse that support it. Captured by the resurrected clone of its wickedest agent and horribly mutilated, only a pair of legendary swords and his native wit and character stand between Sam, his beautiful, mysterious partner and a bloody death!

Poetry

- *Who Shall Sing, When Man is Gone*

  Original poetry, including the epic-length poem about an imagined end of the world brought about by a nuclear war that gives the collection its name. Includes many long and short works on love and life, pain and death.

  > Ocean roaring, whipped by storm
  > in damned defiance, hating hell
  > with every wave and every swell,
  > every shark and every shell
  > and shoreline.

- *Hot Tea!*

  More original poetry with a distinctly Zen cast to it. Works range from funny and satirical to inspiring and uplifting, with a few erotic poems thrown in.

  > Chop water, carry
  > wood. Ice all around,
  > fire is dying. Winter Zen?

All of these books can be found on the online Lulu store here:

http://stores.lulu.com/store.php?fAcctID=877977

The Book of Lilith is available on Amazon, Barnes and Noble and other online bookseller websites.

Contents

Preface

This math text is intended to be used to support the two-semester series of courses teaching *introductory physics* at the college level. Students who hope to succeed in learning physics, from my two online textbooks that teach it or elsewhere, need as a prerequisite a solid grasp of a certain amount of mathematics.

I usually recommend 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) before tackling either semester of physics: Mechanics or Electricity and Magnetism. Students should also have completed single variable *integral* calculus, typified by the BC advanced placement test or a second-semester college calculus course, before taking the second semester course in Electricity and Magnetism. It is usually OK to be taking the second semester course in integral calculus at the same time you are taking the first semester course in physics (Mechanics); that way you are finished in time to start the second semester of physics with all the math you need fresh in your mind.

In my (and most college level) textbooks it is presumed that students are competent in geometry, trigonometry, algebra, and single variable differential and integral 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 of the sort you learn to do in single variable integral calculus. Concepts such as coordinate systems, vectors algebra, the algebra of complex numbers, and at least a couple of series expansions help tremendously – they are taught to some extent in context in the course, but if a student has never seen them before they will probably struggle.

This book (in which you are reading these words) is not really intended to be a "textbook" in math. It is rather a review guide, one that presumes that students have already had a "real" course in most of the math it covers but that perhaps it was some years ago when they took it (or then never did terribly well in it) and need some help relearning the stuff they really, truly need to know in order to be able to learn physics. It is strongly suggested that all physics students that are directed here for help or review *skim read the entire text* right away, reading it just carefully enough that they can see what is there and sort it out into stuff they know and things that maybe they don't know. If you do this well, it won't take very long (a few hours, perhaps a half a day) and afterwords you can use it as a working reference as needed while working on the actual course material.

Introduction

This isn't really a math textbook, but math is an extremely important part of physics. Physics textbooks usually at least attempt to include math support for key ideas, reviewing e.g. how to do a cross product. The problem with this is that this topical review tends to be scattered throughout the text or collected in an appendix that students rarely find when they most need it (either way).

I don't really *like* either of these solutions. This is my own solution to the same problem: a very short *math review* textbook that contains just precisely what is needed in order to really get going with physics in the *introductory* classes one takes as a freshman physics major or later, perhaps as a pre-medical student or math major.

This math is not horrible difficult, but it often (and quite reasonably) is challenging for students of introductory physics. It is often the first time they are called upon to actually *use* a lot of the math they took over several *years* of instruction in high school and college. To my experience, most introductory physics students struggle with *simple* things like decomposing vectors into components and adding them componentwise, with the quadratic formula, with complex numbers, with simple calculus techniques, sometimes even with basic *algebra*.

College level math textbooks tend to be useless to help buff up one's skills in this kind of thing at the level needed to support the physics, alas. One would need a bunch of them – one for vectors, coordinate systems, and trig, one for basic calculus, one to review high school algebra, one for numbers in general and complex numbers in particular, one for basic geometry. It is rare to find a single book that treats all of this and does so *simply and concisely* and without giving the student a dozen examples or exercises per equation or relation covered in the book. What is needed is a *comprehensive review* of material, one that is shallow and fast enough to let a student quickly recall it if they've seen it before well enough to use, yet deep and complete enough that they can get to where they can *work* with the math even if they have *not* had a full course in it, or if they can't remember three words about e.g. complex variables from the two weeks in an algebra class three years ago when they covered them – in high school.

Hence this book. I recommend *skimming it quickly* right now to learn what it contains, then making a slightly slower pass to review it, then go ahead and move on the the physics and come *back* anytime you are stumped by not remembering how to integrate something like (for example):

$$\int_0^\infty x^2 e^{-ax} dx \tag{1}$$

Here are some of the things you should be able to find help for in this book. Note well that this is a work in progress, and not all of them may be in place. Feel free to bug me at rgb at phy dot duke dot edu if something you need or think should be here isn't here. I'm dividing my time between the writing and development of *several* textbooks (including the two semester physics textbooks this short review was once a part of and is now intended to support) but squeaky wheels get the oil, I always say.

This book leverages existing online resources for learning or reviewing math to the extent possible, especially Wikipedia. If you bought a paper copy of this book to help support the author, Thank You! However, I would still recommend that you read through the book in a computer browser from time to time, especially one that supports active links. Most of the footnotes that contain wikipedia pages will pipe you straight through to the referenced pages if clicked!

- **Numbers**

  Integers, real numbers, complex numbers, prime numbers, important numbers, the algebraic representation of numbers. Physics is all about numbers.

- **Algebra**

  Algebra is the symbolic manipulation of numbers according to certain rules to (for example) solve for a particular desired physical quantity in terms of others. We also review various well-known functions and certain expansions.

- **Coordinate Systems and Vectors**

  Cartesian, Cylindrical and Spherical coordinate systems in 2 and 3 dimensions, vectors, vector addition, subtraction, inner (dot) product of vectors, outer (cross) product of vectors.

- **Trigonometric Functions and Complex Exponentials**

  There is a beautiful relationship between the complex numbers and trig functions such as sine, cosine and tangent. This relationship is encoded in the "complex exponential" $e^{i\theta}$, which turns out to be a *very* important and useful relationship. We review this in a way that hopefully will make working with these complex numbers and trig functions both easy.

- **Differentiation**

  We quickly review what differentiation *is*, and then present, sometimes with a quick proof, a table of derivatives of functions that you should know to make learning physics at this level straightforward.

- **Integration**

  Integration is basically antidifferentiation or summation. Since many physical relations involve summing, or integrating, over extended distributions of mass, of charge, of current, of fields, we present a table of integrals (some of them worked out for you in detail so you can see how it goes).

Numbers

Natural, or Counting Numbers

This:

$$1, 2, 3, 4 \dots$$

is the set of numbers[1] that is pretty much the first piece of mathematics most ordinary human beings (and possibly a few extraordinary dogs[2]) learns. They are used to *count*, initially to count things: concrete objects such as pennies or marbles. This is in some respects surprising, since pennies and marbles are never really *identical*, but the mind is very good at classifying things based on their similarities and glossing over their differences. In physics, however, one encounters particles that *are* identical as far as we can tell even with careful observations and measurements – electrons, for example, differ only in their position or orientation.

The natural numbers are usually defined along with a set of operations known as *arithmetic*[3] . The well-known operations of ordinary arithmetic are addition, subtraction, multiplication, and division. One rapidly sees that the set of natural/counting numbers is not *closed* with respect to all of them. That just means that if one subtracts two natural numbers, say 7 from 5, one does not necessarily get a natural number. More concretely, one cannot take seven cows away from a field containing five cows, one cannot remove seven pennies from a row of five pennies.

This helps us understand the term ***closure*** in mathematics. A ***set*** (of, say, numbers) is ***closed with respect to some binary operation*** (say, addition, or subtraction) if any two members of the set combined with the operation produce a member of the set. The natural numbers are closed with respect to addition (the sum of any two natural numbers is a natural number) and multiplication (the product of any two natural numbers is a natural number) but not, if you think about it, subtraction or division. More on this later.

[1]Wikipedia: http://www.wikipedia.org/wiki/Number.
[2]http://www.sciencedaily.com/releases/2009/08/090810025241.htm
[3]Wikipedia: http://www.wikipedia.org/wiki/Arithmetic.

Natural numbers greater than 1 in general can be factored into a representation in *prime numbers*[4] . For example:

$$45 = 2^0 3^2 5^1 7^0 ... \tag{2}$$

or

$$56 = 2^3 3^0 5^0 7^1 11^0 ... \tag{3}$$

This sort of factorization can sometimes be very useful, but not so much in introductory physics.

Infinity

It is easy to see that there is no largest natural number. Suppose there was one, call it L. Now add one to it, forming $M = L + 1$. We know that $L + 1 = M > L$, contradicting our assertion that L was the largest. This *lack* of a largest object, lack of a boundary, lack of termination in series, is of enormous importance in mathematics and physics. If there is no largest number, if there is no "edge" to space or time, then it in some sense they run on *forever*, without termination.

In spite of the fact that there is no *actual* largest natural number, we have learned that it is highly advantageous in many context to invent a *pretend* one. This pretend number doesn't actually exist *as a number*, but rather stands for a certain *reasoning process*.

In fact, there are a number of properties of numbers (and formulas, and integrals) that we can only understand or evaluate if we *imagine* a very large number used as a boundary or limit in some computation, and then let that number mentally increase *without bound*. **Note well** that this is a mental trick, no more, encoding the observation that there is no largest number and so we can increase a number parameter without bound. However, mathematicians and physicists use this mental trick all of the time – it becomes a way for our finite minds to encompass the idea of the infinite, of *unboundedness*. To facilitate this process, we invent a *symbol* for this unreachable limit to the counting process and give it a name.

We call this unboundedness ***infinity***[5] – the lack of a finite boundary – and give it the symbol ∞ in mathematics.

In *many contexts* we will treat ∞ as a number in all of the number systems mentioned below. We will talk blithely about "infinite numbers of digits" in number representations, which means that the digits specifying some number simply keep on going without bound. However, it is ***very important to bear in mind that ∞ is not a number, it is a concept!*** Or at the very least, it is a highly *special* number, one that doesn't satisfy the axioms or participate in the usual operations of ordinary arithmetic. For example, for N

[4]Wikipedia: http://www.wikipedia.org/wiki/Prime Number.
[5]Wikipedia: http://www.wikipedia.org/wiki/Infinity.

any finite number:

$$\begin{aligned}
\infty + \infty &= \infty & (4)\\
\infty + N &= \infty & (5)\\
\infty - \infty &= \text{undefined} & (6)\\
\infty * N &= \infty & (7)\\
N/\infty &= 0 \ (\textbf{\textit{But}}\ 0 * \infty \text{ is not equal to } N!) & (8)\\
\infty/\infty &= \text{ undefined} & (9)
\end{aligned}$$

These are certainly "odd" rules compared to ordinary arithmetic! They all make sense, though, if you replace the symbol with "something (infinitely) bigger than any specific number you can imagine".

For a bit longer than a century now (since Cantor organized set theory and discussed the various ways sets could become infinite and set theory was subsequently axiomatized) there has been an *axiom of infinity* in mathematics postulating its formal existence as a "number" with these and other odd properties.

Our principal use for infinity will be as a limit in calculus and in series expansions. We will use it to describe *both* the very large (but never the largest) *and* reciprocally, the very small (but never quite zero). We will use infinity to name the *process* of taking a small quantity and making it "infinitely small" (but nonzero) – the idea of the *infinitesimal*, or the complementary operation of taking a large (finite) quantity (such as a limit in a finite sum) and making it "infinitely large". These operations do not always make arithmetical sense – consider the infinite sum of the natural numbers, for example – but when they do they are *extremely valuable*.

Integers

To achieve closure in addition, subtraction, and multiplication one introduces negative whole numbers and zero to construct the set of *integers*:

$$..., -3, -2, -1, 0, 1, 2, 3, ...$$

Today we take these things for granted, but in fact the idea of negative numbers in particular is quite recent. Although they were *used* earlier, mathematicians only accepted the idea that negative numbers were legitimate numbers by the latter 19th century! After all, if you are counting cows, how can you add negative cows to an already empty field? How can you remove seven pennies from a row containing only five? Numbers in Western society were thought of as being concrete properties of *things*, tools for bookkeeping, rather than strictly abstract entities about which one could axiomatically reason until well into the Enlightenment[6].

[6]Well, except for the Pythagoreans, who reasoned about – nay, *worshipped* – numbers a couple of thousand years before the Enlightenment. Interested readers might want to look at Morris Kline's

In physics, integers or natural numbers are often represented by the letters i, j, k, l, m, n, although of course in algebra one *does* have a range of choice in letters used, and some of these symbols are "overloaded" (used for more than one thing) in different formulas.

Integers can in general also be factored into primes, but problems begin to emerge when one does this. First, negative integers will always carry a factor of -1 times the prime factorization of its absolute value. But the introduction of a form of "1" into the factorization means that one has to deal with the fact that $-1 * -1 = 1$ and $1 * -1 = -1$. This possibility of permuting negative factors through all of the positive and negative halves of the integers has to be generally ignored because there is a complete symmetry between the positive and negative half-number line; one simply prepends a single -1 to the prime factorization to serve as a reminder of the sign. Second, 0 times anything is 0, so it (and the numbers ± 1) are generally excluded from the factorization process.

Integer arithmetic is associative, commutative, is closed under addition, subtraction and multiplication, and has lots of nice properties you can learn about on e.g. Wikipedia. However, it is still not closed under division! If one divides two integers, one gets a number that is not, in general, an integer!

This forming of the ***ratio*** between two integer quantities leads to the next logical extension of our growing system of numbers: The ***rational numbers***.

Rational Numbers

If one takes two integers a and b and divides a by b to form $\frac{a}{b}$, the result will often *not* be an integer. For example, $1/2$ is not an integer (although $2/1$ is!), nor is $1/3, 1/4, 1/5...$, nor $2/3, 4/(-7) = -4/7, 129/37$ and so on. These numbers are all the *ratios* of two integers and are hence called *rational numbers*[7] .

Rational numbers when expressed in a base[8] e.g. base 10 have an interesting property. Dividing one out produces a finite number of non-repeating digits, followed by a finite sequence of digits that repeats cyclically forever. For example:

$$\frac{1}{3} = 0.3333... \tag{10}$$

or

$$\frac{11}{7} = 1.571428\ 571428\ 571428... \tag{11}$$

(where a small space has been inserted to help you see the pattern).

Mathematics: The Loss of Certainty, a book that tells the rather exciting story of the development of mathematical reasoning from the Greeks to the present, in particular the discovery that mathematical reasoning does not lead to "pure knowledge", *a priori* truth, but rather to contingent knowledge that may or may not apply to or be relevant to the real world.

[7]Wikipedia: http://www.wikipedia.org/wiki/rational number.

[8]The *base* of a number is the range over which each digit position in the number cycles. We generally work and think in base ten, most likely because our ten fingers are among the first things we count! Hence *digit*, which refers to a positional number *or* a finger or toe. However, base two (binary), base eight (octal) and base sixteen (hexadecimal) are all useful in computation, if not physics.

Note that finite precision decimal numbers are precisely those that are terminated with an infinite string of the digit 0, and hence are all rational. That is, if we keep numbers only to the hundredths place, e.g. 4.17, -17.01, 3.14, the assumption is that all the rest of the digits in the number are 0 – 3.14000..., which is rational.

It may not be the case that those digits really *are* zero. We will often be multiplying by $1/3 \approx 0.33$ to get an approximate answer to all of the precision we need in a problem. In any event, we generally *cannot* handle an infinite number of nonzero digits in some base, repeating or not, in our arithmetical operations, so truncated base two or base ten, rational numbers are the special class of numbers over which we do much of our arithmetic, whether it be done with paper and pencil, slide rule, calculator, or computer.

If all rational numbers have digit strings that eventually cyclically repeat, what about all numbers whose digit strings do *not* cyclically repeat? These numbers are *not* rational.

Irrational Numbers

An irrational number[9] is one that *cannot be written* as a ratio of two integers e.g. a/b. It is not immediately obvious that numbers like this exist at all. When rational numbers were discovered (or invented, as you prefer) by the Pythagoreans, they were thought to have nearly mystical properties – the Pythagoreans quite literally worshipped numbers and thought that everything in the Universe could be understood in terms of the ratios of natural numbers. Then *Hippasus*, one of their members, demonstrated that for an isosceles right triangle, if one assumes that the hypotenuse and arm are commensurable (one can be expressed as an integer ratio of the other) that the hypotenuse had to be even, but the legs had to be both even and odd, a contradiction. Consequently, it was certain that they could *not* be placed in a commensurable ratio – the lengths are related by an *irrational* number.

According to the (possibly apocryphal) story, Hippasus made this discovery on a long sea voyage accompanied by a group of fellow Pythagoreans, and they were so annoyed at his *blasphemous* discovery that their religious beliefs in the rationality of the Universe (so to speak) were false that they *threw him overboard* to drown! Folks took their mathematics quite seriously back then!

As we've seen, all digital representation of finite precision or digital representations where the digits eventually cycle correspond to rational numbers. Consequently its digits in a decimal representation of an irrational number *never* reach a point where they cyclically repeat or truncate (are terminated by an infinite sequence of 0's).

Many numbers that are of great importance in physics, especially $e = 2.718281828...$ and $\pi = 3.141592654...$ are irrational, and we'll spend some time discussing both of them below. When working in coordinate systems, many of the trigonometric ratios for "simple" right triangles (such as an isoceles right triangle) involve numbers such as $\sqrt{2}$, which are also irrational – this was the basis for the earliest proofs of the existence of

[9] Wikipedia: http://www.wikipedia.org/wiki/irrational number.

irrational numbers, and $\sqrt{2}$ was arguably the first irrational number discovered.

Whenever we compute a number answer we *must* use rational numbers to do it, most generally a finite-precision decimal representation. For example, 3.14159 may *look* like π, an irrational number, but it is really $\frac{314159}{100000}$, a rational number that *approximates* π to six significant figures.

Because we *cannot* precisely represent them in digital form, in physics (and mathematics and other disciplines where precision matters) we will often carry important irrationals along with us in computations as *symbols* and only evaluate them numerically at the end. It is important to do this because we work quite often with functions that yield a rational number or even an integer when an irrational number is used as an argument, e.g. $\cos(\pi) = -1$. If we did finite-precision arithmetic prematurely (on computer or calculator) we might well end up with an *approximation* of -1, such as -0.999998 and could not be sure if it was *supposed* to be -1 or really was supposed to be a bit more.

There are lots of nifty truths regarding rational and irrational numbers. For example, in between any two rational numbers lie an *infinite* number of *irrational* numbers. This is a "bigger infinity"[10] than *just* the countably infinite number of integers or rational numbers, which actually has some important consequences in physics – it is one of the origins of the theory of deterministic chaos.

Real Numbers

The union of the irrational and rational numbers forms the *real number line*.[11] Real numbers are of great importance in physics. They are closed under the arithmetical operations of addition, subtraction, multiplication and division, where one must exclude only division by zero[12]. Real exponential functions such as a^b or e^x (where a, b, e, x are all presumed to be real) will have real values, as will algebraic functions such as $(a + b)^n$ where n is an integer.

However, as before we can discover arithmetical operations, such as the power operation (for example, the square root, $\sqrt{x} = x^{1/2}$ for some real number x) that lead to problems with closure. For positive real arguments $x \geq 0$, $y = \sqrt{x}$ is real, but probably irrational (irrational for most possible values of x). But what happens when we try to form the square root of negative real numbers? In fact, what happens when we try to form the square root of -1?

This is a bit of a problem. All real numbers, squared or taken to any *even* integer

[10]Wikipedia: http://www.wikipedia.org/wiki/infinity. There are (at least) two different kinds of infinity – countable and uncountable. Countable doesn't mean that one can count to infinity – it means one can create a one-to-one map between the (countably infinite) counting numbers and the countably infinite set in question. Uncountable means that one cannot make this mapping. The set of all real numbers in any finite interval form a *continuum* and is an example of an uncountably infinite set.

[11]Wikipedia: http://www.wikipedia.org/wiki/real line.

[12]Which we will often take in the sense of division by a small number ***in the limit*** that the small number goes to zero, which is one of several ways to reach/define ∞. In certain contexts, then, division by zero can be considered to be "∞", where in others it is simply undefined.

power, are positive. There therefore *is no* real number that can be squared to make -1. All we can do is *imagine* such a number, i, and then make our system of numbers bigger still to accomodate it. This process leads us to the *imaginary* unit i such that $i^2 = -1$, to all possible products and sums of this number and our already known real numbers and thereby to numbers with both ***real*** (no necessary factor of i) and ***imaginary*** (a necessary factor of i) parts. Such a number might be represented in terms of real numbers like:

$$z = x + iy \tag{12}$$

where x and y are plain old real numbers and i is the imaginary unit.

Whew! A number that is now the sum of two very different *kinds* of number. That's complicated! Let's call these new beasts ***complex numbers***.

Complex Numbers

At this point you should begin to have the feeling that this process of generating supersets of the numbers we already have figured out that close under additional operations or have some desired additional properties will never end. You would be right, and some of the extensions (division algebras that we will not cover here such as quaternions[13] or more generally, geometric algebras[14]) are actually very useful in more advanced physics. However, we have a finite amount of time to *review* numbers here, and complex numbers are the most we will need in *this* course (or "most" undergraduate physics courses even at a somewhat more advanced level). They are important enough that we'll spend a whole section discussing them below; for the moment we'll just define them.

We start with the unit imaginary number[15] , i. You *might* be familiar with the *naive* definition of i as the square root of -1:

$$i = +\sqrt{-1} \tag{13}$$

This definition is common but slightly unfortunate. If we adopt it, we have to be careful *using* this definition in algebra or we will end up proving any of the many variants of the following:

$$-1 = i \cdot i = \sqrt{-1} \cdot \sqrt{-1} = \sqrt{-1 \cdot -1} = \sqrt{1} = 1 \tag{14}$$

Oops.

A better definition for i that it is just the algebraic number such that:

$$i^2 = -1 \tag{15}$$

[13]Wikipedia: http://www.wikipedia.org/wiki/Quaternions.
[14]Wikipedia: http://www.wikipedia.org/wiki/Geometric Algebra.
[15]Wikipedia: http://www.wikipedia.org/wiki/imaginary unit.

and to leave the square root bit out. Thus we have the following ***well-defined*** cycle:

$$
\begin{aligned}
i^0 &= 1 \\
i^1 &= i \\
i^2 &= -1 \\
i^3 &= (i^2)i = -1 \cdot i = -i \\
i^4 &= (i^2)(i^2) = -1 \cdot -1 = 1 \\
i^5 &= (i^4)i = i \\
&...
\end{aligned} \tag{16}
$$

where we can use these rules to do the following sort of simplification:

$$+\sqrt{-\pi b} = +\sqrt{i^2 \pi b} = +i\sqrt{\pi b} \tag{17}$$

but where we never actually write $i = \sqrt{-1}$.

We can make all the ***imaginary numbers*** by simply scaling i with a real number. For example, $14i$ is a purely imaginary number of magnitude 14. $i\pi$ is a purely imaginary number of magnitude π. All the purely imaginary numbers therefore form the ***imaginary line*** that is basically the real line, times i. Note well that this line ***contains the real number zero*** – 0 is in fact the *intersection* of the imaginary line and the real line.

With this definition, we can define an arbitrary ***complex number z*** as the sum of an arbitrary real number plus an arbitrary imaginary number:

$$z = x + iy \tag{18}$$

where x and y are both real numbers. It can be shown that the roots of any polynomial function can always be written as complex numbers, making complex numbers of great importance in physics. However, their *real* power in physics comes from their relation to exponential functions and trigonometric functions.

Complex numbers (like real numbers) form a *division algebra*[16] – that is, they are closed under addition, subtraction, multiplication, *and division.* Division algebras permit the factorization of expressions, something that is obviously very important if you wish to algebraically solve for quantities.

Hmmmm, seems like we ought to look at this "algebra" thing. Just what *is* an algebra? How does algebra work?

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

Algebra

Algebra[17] is a *reasoning process* that is one of the fundamental cornerstones of mathematical reasoning. As far as we are concerned, it consists of two things:

- Representing *numbers* of any of the types discussed above (where we might as well assume that they are *complex numbers* since real numbers are complex, rational and irrational numbers are real, integers are rational, and natural numbers are integers, so natural numbers and all of the rest of them are also complex) with ***symbols***. In physics this representation isn't only a matter of knowns and unknowns – we will often use algebraic symbols for numbers we know or for parameters in problems even when their value is actually given as part of the problem. In fact, with only a relatively few exceptions, we will prefer to use symbolic algebra ***as much as we can*** to permit our algebraic manipulations to *eliminate as much eventual arithmetic* (computation involving actual numbers) from a problem as possible.
- Performing a sequence of *algebraic transformations* of a set of symbolic equations or inequalities to convert it from one (uninformative) form to another (desired, informative) form. These transformations are generally based on the set of arithmetic operations defined (and permitted!) over the field(s) of the number type(s) being manipulated.

That's it.

Note well that it isn't ***always*** a matter of solving for some unknown but determined variable in terms of known variables, although this is certainly a useful thing to be able to do. Algebra is just as often used to derive ***relationships*** and hence gain ***conceptual insight*** into a system being studied, possibly expressed as a derived ***law***. Algebra is thus in some sense the ***conceptual language*** of physics as well as the set of tools we use to solve problems within the context of that language. English (or other spoken/written human languages) is too imprecise, too multivalent, too illogical and inconsistent to serve as a good language for this purpose, but algebra (and related ***geometries***) are just perfect.

The transformations of algebra applied to equalities (the most common case) can be summarized as follows (non-exhaustively). If one is given one or more equations involving

[17] Wikipedia: http://www.wikipedia.org/wiki/algebra.

a set of variables $a, b, c, ...x, y, z$ one can:

1. Add any scalar number or well defined and consistent symbol to both sides of any equation. Note that in physics problems, symbols carry units and it is necessary to add only symbols that *have the same units* as we cannot, for example, add seconds to kilograms and end up with a result that makes any sense!

2. Subtract any scalar number or consistent symbol ditto. This isn't really a separate rule, as subtraction is just adding a negative quantity.

3. Multiplying both sides of an equation by any scalar number or consistent symbol. In physics one *can* multiply symbols with different units, such an equation with (net) units of meters times a symbol given in seconds.

4. Dividing both sides of an equation ditto, save that one has to be careful when performing symbolic divisions to avoid points where division is not permitted or defined (e.g. dividing by zero or a variable that might take on the value of zero). Note that dividing one unit by another in physics is also permitted, so that one can sensibly divide length in meters by time in seconds.

5. Taking both sides of an equation to any power. Again some care must be exercised, especially if the equation can take on negative or complex values or has any sort of domain restrictions. For fractional powers, one may well have to specify the *branch* of the result (which of many possible roots one intends to use) as well.

6. Placing the two sides of any equality into *almost* any functional or algebraic form, either given or known, as if they are variables of that function. Here there are some serious caveats in both math and physics. In physics, the most important one is that if the functional form has a power-series expansion then the equality one substitutes in must be *dimensionless*. This is easy to understand. Supposed I know that x is a length in meters. I *could* try to form the exponential of x: e^x, but if I expand this expression, $e^x = 1 + x + x^2/2! + ...$ which is *nonsense!* How can I add meters to meters-squared? I can only exponentiate x if it is dimensionless. In mathematics one has to worry about the domain and range. Suppose I have the relation $y = 2 + x^2$ where x is a real (dimensionless) expression, and I wish to take the $\cos^{-1}$ of both sides. Well, the *range* of cosine is only -1 to 1, and my function y is clearly strictly larger than 2 and cannot have an inverse cosine! This is obviously a powerful, but *dangerous* tool.

In the sections below, we'll give examples of each of these points and demonstrate some of the key algebraic methods used in physics problems.

0.1 Symbols and Units

In the first chapter, we went over ***numbers*** of various sorts. We stopped short of fully deriving the rules of arithmetic, and in this course we will assume – for better or worse

– that ***the reader can do basic arithmetic without using a calculator***. That is, if you cannot yet add, subtract, multiply and divide ordinary (e.g. integer, or rational) numbers, it is time for you to ***back up and start acquiring these basic life skills*** before tackling a physics course. A calculator doing ill-understood magic is no substitute for an educated human brain.

Similarly, if you don't know what square roots (or cube roots or nth roots) or general powers of numbers are, and can't find or recognize (for example) the square root of 25 or the cube root of 27 when asked, if you don't know (or cannot easily compute) that $2^5 = 32$, that $10^3 = 1000$, that $5^{3/2} = 5\sqrt{5}$, it's going to be very difficult to follow algebra that requires you to take symbols that stand for these quantities and abstract them as $\sqrt{a}$, or $b^{3/2}$.

The good news, though, is that the whole point of algebra is to *divorce* doing arithmetic to the greatest extent possible from the process of *mathematical reasoning* needed to solve any given problem. Most physicists, myself included, hate doing boring arithmetic as much as you do. That's why we don't spend much time doing it – that's what abaci and slide rules (just kidding), calculators, and computers are for. But no arithmetic-doing machine will give you the right answer to a complex problem unless you have *solved* the problem so that you know exactly what to ask it to do.

Enter algebra.

Suppose I want to add two numbers, but I don't know what they are yet. I know, this sounds a bit bizarre – how can I know I'll need to add them without knowing what they are? If you think about it, though, this sort of thing *happens all of the time!* A relay race is run, and you are a coach timing the splits at practice. You click a stop watch when each baton is passed and when the final runner crosses the finish line. You are going to want to know the total time for the race, but used up all of your stop watches timing the splits. So you are going to need to *add the four times* even though you don't yet know what they are.

Except that *you* don't want to do the addition. That's what junior assistant coaches are for. So you give the junior assistant coach a set of instructions. Verbally, they are "Take the time for the first runner, the second runner, the third runner, and the fourth runner and add them to find the total time for the relay team in the practice race".

But that's awfully wordy. So you invent ***symbols*** for the four times that need to be summed. Something pithy, yet informative. It can't be too complex, as your junior assistant coach isn't too bright. So you use (say) the symbol 't' for time, and use an *index* to indicate the runner/stop watch. The splits become: t_1, t_2, t_3, t_4. Your rule becomes:

$$t_{\rm tot} = t_1 + t_2 + t_3 + t_4 \tag{19}$$

But that is *still* a bit long, and what happens if you want to form the average of all of the splits run during a day? There might be 40 or 50 of them! That's way to tedious to write out. So you invent *another* symbol, one that stands for ***summation***. Summation begins with an 'S', but if we use English alphabet symbols for operations, we risk confusing people if we also use them for quantities. So we pick a nice, safe *Greek* letter S, the ***capital sigma***: Σ:

$$t_{\mathrm{tot}} = \sum_{i=1}^{4} t_i = t_1 + t_2 + t_3 + t_4 \tag{20}$$

Note that we've invented/introduced *another* symbol, i. i stands for the ***index*** of the time being added. We decorate the Σ in the summation symbol with the ***limits*** of the sum. We want to start with the first ($i = 1$) time and end the repetitive summing with the last ($i = 4$) time. Now we can easily change our instructions to the junior assistant coach:

$$t_{\mathrm{tot}} = \sum_{i=1}^{52} t_i = t_1 + t_2 + ... + t_{52} \tag{21}$$

for the four splits each in each of the thirteen races being run in the relay.

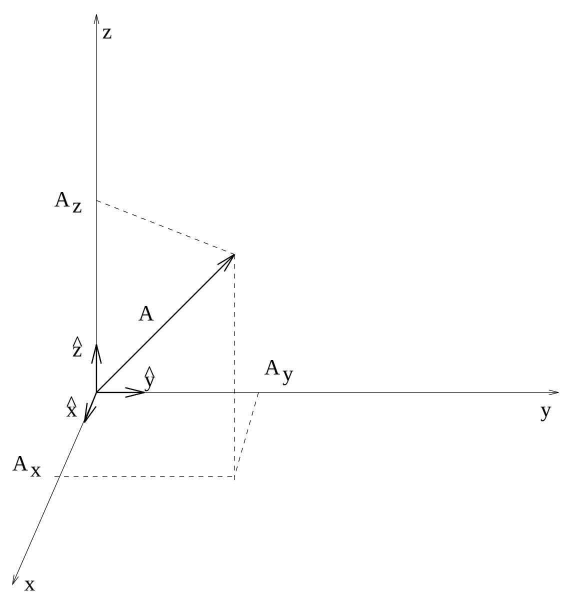
OK, so at this point we'll abandon the coaching metaphor, but the point should be clear. Accountants need ways to represent adding up your bank accounts, even though they don't know how many accounts you have coming into the office. The army needs ways of representing how much food to get for N soldiers who will spend D days in the field, each of them eating M meals per day – and at the end needs to be able to compute the total cost of providing it, given a cost per meal of Q. At some point, algebra becomes more than a vehicle for defining arithmetic that needs to be done – it becomes a way of ***reasoning*** about who ***systems*** of things – like that last example. One can reduce the army's food budget for a campaign to an *algebraic form* for the cost C:

$$C = N * D * M * Q \tag{22}$$

THat's a lot shorter, and easier to check, than the paragraphs of text that would be needed to describe it.

Coordinate Systems, Points, Vectors

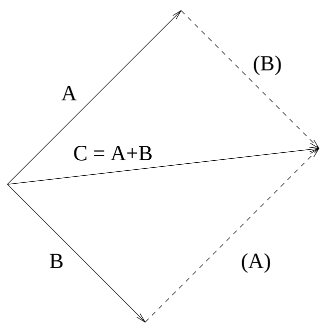
Review of Vectors

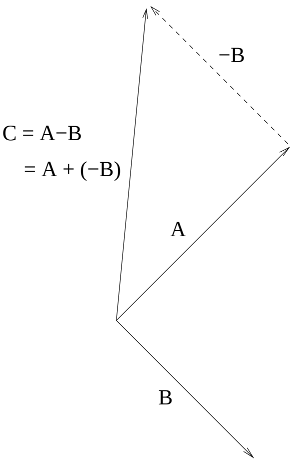


(23)

Most motion is not along a straight line. If fact, almost no motion is along a line. We therefore need to be able to describe motion along *multiple dimensions* (usually 2 or 3). That is, we need to be able to consider and evaluate *vector* trajectories, velocities, and accelerations. To do this, we must first learn about what vectors are, how to add, subtract or decompose a given vector in its cartesian coordinates (or equivalently how to convert between the cartesian, polar/cylindrical, and spherical coordinate systems), and what scalars are. We will also learn a couple of products that can be constructed from vectors.

A **vector** in a coordinate system is a directed line between two points. It has **mag-**





nitude and **direction**. Once we define a coordinate origin, each particle in a system has a **position vector** (e.g. – $\vec{A}$) associated with its location in space drawn from the origin to the physical coordinates of the particle (e.g. – (A_x, A_y, A_z)):

$$\vec{A} = A_x\hat{x} + A_y\hat{y} + A_z\hat{z} \tag{24}$$

The position vectors clearly depend on the choice of coordinate origin. However, the **difference vector** or **displacement vector** between two position vectors does **not** depend on the coordinate origin. To see this, let us consider the **addition** of two vectors:

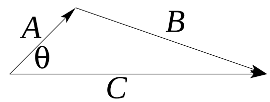
$$\vec{A} + \vec{B} = \vec{C} \tag{25}$$

Note that vector addition proceeds by putting the tail of one at the head of the other, and constructing the vector that completes the triangle. To numerically evaluate the sum of two vectors, we determine their components and add them componentwise, and then reconstruct the total vector:

$$C_x = A_x + B_x \tag{26}$$
$$C_y = A_y + B_y \tag{27}$$
$$C_z = A_z + B_z \tag{28}$$

If we are given a vector in terms of its **length** (magnitude) and **orientation** (direction angle(s)) then we must evaluate its cartesian components before we can add them (for example, in 2D):

$$A_x = \left|\vec{A}\right| \cos(\theta_A) \qquad B_x = \left|\vec{B}\right| \cos\theta_B \tag{29}$$
$$A_y = \left|\vec{A}\right| \sin(\theta_A) \qquad B_y = \left|\vec{B}\right| \sin\theta_B \tag{30}$$



This process is called **decomposing** the vector into its cartesian components.

The **difference** between two vectors is defined by the addition law. Subtraction is just adding the negative of the vector in question, that is, the vector with the **same** magnitude but the **opposite** direction. This is consistent with the notion of adding or subtracting its components. Note well: Although the vectors themselves may depend upon coordinate system, the difference between two vectors (also called the **displacement** if the two vectors are, for example, the postion vectors of some particle evaluated at two different times) does **not**.

When we reconstruct a vector from its components, we are just using the law of vector addition itself, by **scaling** some special vectors called **unit vectors** and then adding them. Unit vectors are (typically perpendicular) vectors that define the essential directions and orientations of a coordinate system and have unit length. Scaling them involves multiplying these unit vectors by a number that represents the magnitude of the vector component. This scaling number has no direction and is called a **scalar**. Note that the product of a vector and a scalar is always a vector:

$$\vec{B} = C\vec{A} \tag{31}$$

where C is a scalar (number) and $\vec{A}$ is a vector. In this case, $\vec{A} \,||\, \vec{B}$ ($\vec{A}$ is parallel to $\vec{B}$).

In addition to multiplying a scalar and a vector together, we can define products that multiply two vectors together. By "multiply" we mean that if we double the magnitude of either vector, we double the resulting product – the product is *proportional* to the magnitude of either vector. There are two such products for the ordinary vectors we use in this course, and both play *extremely important roles* in physics.

The first product creates a scalar (ordinary number with magnitude but no direction) out of two vectors and is therefore called a **scalar product** or (because of the multiplication symbol chosen) a **dot product**. A scalar is often thought of as being a "length" (magnitude) on a single line. Multiplying two scalars on that line creates a number that has the *units* of length squared but is geometrically not an area. By selecting as a direction for that line the direction of the vector itself, we can use the scalar product to *define* the length of a vector as the *square root* of the vector magnitude times itself:

$$\left|\vec{A}\right| = +\sqrt{\vec{A}\cdot\vec{A}} \tag{32}$$

From this usage it is clear that a scalar product of two vectors can never be thought of as an area. If we generalize this idea (preserving the need for our product to be symmetrically proportional to both vectors, we obtain the following definition for the general scalar product:

$$\vec{A}\cdot\vec{B} = A_x * B_x + A_y * B_y \ldots \tag{33}$$

$$= \left|\vec{A}\right|\left|\vec{B}\right|\cos(\theta_{AB}) \tag{34}$$

This definition can be put into *words* – a scalar product is the length of one vector (either one, say $|\vec{A}|$) times the *component* of the other vector ($|\vec{B}|\cos(\theta_{AB})$ that points in the *same direction* as the vector $\vec{A}$. Alternatively it is the length $|\vec{B}|$ times the component of $\vec{A}$ parallel to $\vec{B}$, $|\vec{A}|\cos(\theta_{AB})$. This product is *symmetric* and *commutative* ($\vec{A}$ and $\vec{B}$ can appear in either order or role).

The other product multiplies two vectors in a way that creates a third vector. It is called a **vector product** or (because of the multiplication symbol chosen) a **cross product**. Because a vector has magnitude and direction, we have to specify the product in such a way that both are defined, which makes the cross product more complicated than the dot product.

As far as magnitude is concerned, we already used the non-areal combination of vectors in the scalar product, so what is left is the product of two vectors that makes an *area* and not just a "scalar length squared". The area of the parallelogram defined by two vectors is just:

$$\text{Area in } \tilde{\text{A}} \times \tilde{\text{B}} \text{ parallelogram} = \left|\vec{A}\right| \left|\vec{B}\right| sin(\theta_{AB}) \tag{35}$$

which we can interpret as "the magnitude of $\vec{A}$ times the component of $\vec{B}$ perpendicular to $\vec{A}$" or vice versa. Let us accept this as the magnitude of the cross product (since it clearly has the proportional property required) and look at the direction.

The area is nonzero only if the two vectors do *not* point along the same line. Since two non-colinear vectors always lie in (or define) a plane (in which the area of the parallelogram itself lies), and since we want the resulting product to be independent of the coordinate system used, one sensible direction available for the product is along the line *perpendicular to this plane*. This still leaves us with *two* possible directions, though, as the plane has two sides. We have to pick one of the two possibilities by *convention* so that we can communicate with people far away, who might otherwise use a counterclockwise convention to build screws when we used a clockwise convention to order them, whereupon they send us left handed screws for our right handed holes and everybody gets all irritated and everything.

We therefore *define* the direction of the cross product using the *right hand rule*:

> Let the fingers of your *right hand* lie along the direction of the first vector in a cross product (say $\vec{A}$ below). Let them curl naturally through the *small angle* (observe that there are two, one of which is larger than π and one of which is less than π) into the direction of $\vec{B}$. The erect *thumb* of your right hand then points in the general direction of the cross product vector – it at least indicates which of the two perpendicular lines should be used as a direction, unless your thumb and fingers are all double jointed or your bones are missing or you used your left-handed right hand or something.

Putting this all together mathematically, one can show that the following are two equivalent ways to write the cross product of two three dimensional vectors. In components:

$$\vec{A} \times \vec{B} = (A_x * B_y - A_y * B_x)\hat{z} + (A_y * B_z - A_z * B_y)\hat{x} + (A_z * B_x - A_x * B_z)\hat{y} \quad (36)$$

where you should note that x, y, z appear in *cyclic order* (xyz, yzx, zxy) in the positive terms and have a minus sign when the order is *anticyclic* (zyx, yxz, xzy). The product is *antisymmetric* and *non-commutative.* In particular

$$\vec{A} \times \vec{B} = -\vec{B} \times \vec{A} \quad (37)$$

or the product *changes sign* when the order of the vectors is reversed.

Alternatively in *many* problems it is easier to just use the form:

$$\left|\vec{A} \times \vec{B}\right| = \left|\vec{A}\right| \left|\vec{B}\right| sin(\theta_{AB}) \quad (38)$$

to compute the magnitude and assign the direction *literally* by (right) "hand", along the right-handed normal to the AB plane according to the right-hand rule above.

Note that this *axial* property of cross products is realized in nature by things that *twist* or *rotate around an axis.* A screw advances into wood when twisted clockwise, and comes out of wood when twisted counterclockwise. If you let the fingers of your right hand curl around the screw *in the direction of the twist* your *thumb* points in the direction the screw moves, whether it is in or out of the wood. Screws are therefore by convention *right handed.*

One final remark before leaving vector products. We noted above that scalar products and vector products are closely connected to the notions of *length* and *area*, but mathematics per se need not specify the *units* of the quantities multiplied in a product (that is the province of physics, as we shall see). We have numerous examples where two *different* kinds of vectors (with different units but referred to a common coordinate system for direction) are multiplied together with one or the other of these products. In actual fact, there often *is* a buried squared length or area (which we now agree are different kinds of numbers) in those products, but it won't always be obvious in the dimensions of the result.

Two of the most important uses of the scalar and vector product are to define the *work* done as the force through a distance (using a scalar product as work is a scalar quantity) and the *torque* exerted by a force applied at some distance from a center of rotation (using a vector product as torque is an axial vector). These two quantities (work and torque) have the *same units* and yet are very *different* kinds of things. This is just one example of the ways geometry, algebra, and units all get mixed together in physics.

At first this will be very confusing, but remember, back when you where in third grade multiplying *integer numbers* was very confusing and yet rational numbers, irrational numbers, general real numbers, and even complex numbers were all waiting in the wings. This is more of the same, but all of the additions will *mean something* and have a compelling *beauty* that comes out as you study them. Eventually it all makes very, very good sense.

Functions

One of the most important concepts in algebra is that of the *function*. The formal mathematical definition of the term function[18] is beyond the scope of this short review, but the summary below should be more than enough to work with.

A function is a *mapping* between a set of *coordinates* (which is why we put this section *after* the section on coordinates) and a *single value*. Note well that the "coordinates" in question do *not have to be space and/or time*, they can be any set of *parameters* that are relevant to a problem. In physics, coordinates can be any or all of:

- Spatial coordinates, x, y, z
- Time t
- Momentum p_x, p_y, p_z
- Mass m
- Charge q
- Angular momentum, spin, energy, isospin, flavor, color, and much more, including "spatial" coordinates we *cannot see* in exotica such as string theories or supersymmetric theories.

Note well that many of these things can *equally* well be functions themselves – a potential energy function, for example, will usually return the value of the potential *energy* as a function of some mix of spatial coordinates, mass, charge, and time. Note that the coordinates can be continuous (as most of the ones above are classically) or *discrete* – charge, for example, comes only multiples of e and color can only take on three values.

One formally denotes functions in the notation e.g. $F(\vec{x})$ where F is the *function name* represented symbolically and $\vec{x}$ is the entire vector of coordinates of all sorts. In physics we often learn or derive functional forms for important quantities, and may or may not express them as functions in this form. For example, the kinetic energy of a

[18] Wikipedia: http://www.wikipedia.org/wiki/Function (mathematics).

particle can be written either of the two following ways:

$$K(m, \vec{\boldsymbol{v}}) = \frac{1}{2}mv^2 \tag{39}$$

$$K = \frac{1}{2}mv^2 \tag{40}$$

These two forms are equivalent in physics, where it is usually "obvious" (at least when a student has studied adequately and accumulated some practical experience solving problems) when we write an expression just what the variable parameters are. Note well that we not infrequently use *non*-variable parameters – in particular constants of nature – in our algebraic expressions in physics as well, so that:

$$U = -\frac{Gm_1m_2}{r} \tag{41}$$

is a function of m_1, m_2, and r but includes the gravitational constant $G = 6.67 \times 10^{-11}$ N-m^2/kg^2 in symbolic form. Not all symbols in physics expressions are variable parameters, in other words.

One important property of the mapping required for something to be a true "function" is that there must be only a *single value* of the function for any given set of the coordinates. Two other important definitions are:

Domain The *domain* of a function is the set of all of the coordinates of the function that give rise to unique non-infinite values for the function. That is, for function $f(x)$ it is all of the x's for which f is well defined.

Range The *range* of a function is the set of all values of the function f that arise when its coordinates vary across the entire domain.

For example, for the function $f(x) = \sin(x)$, the domain is the entire real line $x \in (-\infty, \infty)$ and the range is $f \in [-1, 1]$[19].

Two last ideas that are of great use in solving physics problems algebraically are the notion of *composition* of functions and the *inverse* of a function.

Suppose you are given two functions: one for the potential energy of a mass on a spring:

$$U(x) = \frac{1}{2}kx^2 \tag{42}$$

where x is the distance of the mass from its equilibrium position and:

$$x(t) = x_0 \cos(\omega t) \tag{43}$$

which is the position as a function of time. We can form the composition of these two functions by substituting the second into the first to obtain:

$$U(t) = \frac{1}{2}kx_0^2 \cos^2(\omega t) \tag{44}$$

[19] This should be read as "x is contained in the open set of real numbers from minus infinity to infinity" (open because infinity is not a number and hence cannot be in the set) and "f is contained in the closed set of real numbers from minus one to one" (closed because the numbers minus one and one are in the set).

This sort of "substitution operation" (which we will rarely refer to by name) is an *extremely important* part of solving problems in physics, so keep it in mind at all times!

With the composition operation in mind, we can define the inverse. Not all functions have a unique inverse function, as we shall see, but most of them have an inverse function that we can use *with some restrictions* to solve problems.

Given a function $f(x)$, *if* every value in the range of f corresponds to one and only one value in its domain x, then $f^{-1} = x(f)$ is *also* a function, called the *inverse* of f. When this condition is satisfied, the range of $f(x)$ is the domain of $x(f)$ and vice versa. In terms of composition:

$$x_0 = x(f(x_0)) \tag{45}$$

and

$$f_0 = f(x(f_0)) \tag{46}$$

for any x_0 in the domain of $f(x)$ and f_0 in the range of $f(x)$ are *both true*; the composition of f and the inverse function for some value f_0 yields f_0 again and is hence an "identity" operation on the range of $f(x)$.

Many functions do not *have* a unique inverse, however. For example, the function:

$$f(x) = \cos(x) \tag{47}$$

does not. If we look for values x_m in the domain of this function such that $f(x_m) = 1$, we find an *infinite number*:

$$x_m = 2\pi m \tag{48}$$

for $m = 0, \pm 1, \pm 2, \pm 3...$ The mapping is then one value in the range to *many* in the domain and the inverse of $f(x)$ is not a *function* (although we can still write down an expression for all of the values that each point in the range maps into when inverted).

We can get around this problem by *restricting the domain* to a region where the inverse mapping *is* unique. In this particular case, we can define a function $g(x) = \sin^{-1}(x)$ where the domain of g is only $x \in [-1, 1]$ and the *range* of g is restricted to be $g \in [-\pi/2, \pi/2)$. If this is done, then $x = f(g(x))$ for all $x \in [-1, 1]$ and $x = g(f(x))$ for all $x \in [-\pi/2, \pi/2)$. The inverse function for many of the functions of interest in physics have these sorts of restrictions on the range and domain in order to make the problem well-defined, and in many cases we have some degree of *choice* in the best definition for any given problem, for example, we could use *any* domain of width π that begins or ends on an odd half-integral multiple of π, say $x \in (\pi/2, 3\pi/2]$ or $x \in [9\pi/2, 11\pi/2)$ if it suited the needs of our problem to do so when computing the inverse of $\sin(x)$ (or similar but different ranges for $\cos(x)$ or $\tan(x)$) in physics.

In a related vein, if we examine:

$$f(x) = x^2 \tag{49}$$

and try to construct an inverse function we discover two interesting things. First, there are two values in the domain that correspond to each value in the range because:

$$f(x) = f(-x) \tag{50}$$

for all x. This causes us to define the inverse function:

$$g(x) = \pm x^{1/2} = \pm\sqrt{x} \tag{51}$$

where the sign in this expression selects one of the two possibilities.

The second is that once we have defined the inverse functions for either trig functions or the quadratic function in this way so that they have restricted domains, it is natural to ask: Do these functions have any meaning for the *unrestricted* domain? In other words, if we have defined:

$$g(x) = +\sqrt{x} \tag{52}$$

for $x \geq 0$, does $g(x)$ exist for *all* x? And if so, what kind of number is g?

This leads us naturally enough into our next section (so keep it in mind) but first we have to deal with several important ideas.

Polynomial Functions

A polynomial function is a sum of monomials:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_n x^n + \ldots \tag{53}$$

The numbers $a_0, a_1, \ldots, a_n, \ldots$ are called the *coefficients* of the polynomial.

This sum can be finite and terminate at some n (called the *degree* of the polynomial) or can (for certain series of coefficients with "nice" properties) be infinite and converge to a well defined function value. Everybody should be familiar with at least the following forms:

$$\begin{aligned} f(x) &= a_0 \quad (\text{0th degree, constant}) & (54)\\ f(x) &= a_0 + a_1 x \quad (\text{1st degree, linear}) & (55)\\ f(x) &= a_0 + a_1 x + a_2 x^2 \quad (\text{2nd degree, quadratic}) & (56)\\ f(x) &= a_0 + a_1 x + a_2 x^2 + a_3 x^3 \quad (\text{3rd degree, cubic}) & (57) \end{aligned}$$

where the first form is clearly independent of x altogether.

Polynomial functions are a simple key to a huge amount of mathematics. For example, differential calculus. It is easy to derive:

$$\frac{dx^n}{dx} = n x^{n-1} \tag{58}$$

It is similarly simple to derive

$$\int x^n dx = \frac{1}{n+1} x^{n+1} + \text{constant} \tag{59}$$

and we will derive both below to illustrate methodology and help students remember these two *fundamental* rules.

Next we note that many continuous functions can be defined in terms of their *power series* expansion. In fact *any* continuous function can be expanded in the vicinity of a point as a power series, and many of our favorite functions have well known power series that serve as an alternative definition of the function. Although we will not derive it here, one extremely general and powerful way to *compute* this expansion is via the *Taylor series*. Let us define the Taylor series and its close friend and companion, the binomial expansion.

The Taylor Series and Binomial Expansion

Suppose $f(x)$ is a continuous and infinitely differentiable function. Let $x = x_0 + \Delta x$ for some Δx that is "small". Then the following is true:

$$\begin{aligned} f(x_0 + \Delta x) &= f(x)\Big|_{x=x_0} + \frac{df}{dx}\Big|_{x=x_0} \Delta x + \frac{1}{2!}\frac{d^2 f}{dx^2}\Big|_{x=x_0} \Delta x^2 \\ &\quad + \frac{1}{3!}\frac{d^3 f}{dx^3}\Big|_{x=x_0} \Delta x^3 + \ldots \end{aligned} \tag{60}$$

This sum will always converge to the function value (for smooth functions and small enough Δx) if carried out to a high enough degree. Note well that the Taylor series can be rearranged to become the *definition* of the derivative of a function:

$$\frac{df}{dx}\Big|_{x=x_0} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} + \mathcal{O}(\Delta x) \tag{61}$$

where the latter symbols stands for "terms of order Δx or smaller" and vanishes in the limit. It can similarly be rearranged to form formal definitions for the second or higher order derivatives of a function, which turns out to be very useful in computational mathematics and physics.

We will find many uses for the Taylor series as we learn physics, because we will *frequently* be interested in the value of a function "near" some known value, or in the limit of very large or very small arguments. Note well that the Taylor series expansion for any polynomial *is* that polynomial, possibly re-expressed around the new "origin" represented by x_0.

To this end we will find it *very* convenient to define the following *binomial expansion*. Suppose we have a function that can be written in the form:

$$f(x) = (c + x)^n \tag{62}$$

where n can be *any real or complex number*. We'd like expand this using the Taylor series in terms of a "small" parameter. We therefore factor out the *larger* of x and c from this expression. Suppose it is c. Then:

$$f(x) = (c + x)^n = c^n (1 + \frac{x}{c})^n \tag{63}$$

where $x/c < 1$. x/c is now a suitable "small parameter" and we can expand this expression around $x = 0$:

$$f(x) = c^n \left(1 + n\frac{x}{c} + \frac{1}{2!}n(n-1)\left(\frac{x}{c}\right)^2 + \frac{1}{3!}n(n-1)(n-2)\left(\frac{x}{c}\right)^3 + \dots\right) \quad (64)$$

Evaluate the derivatives of a Taylor series around $x = 0$ to verify this expansion. Similarly, if x were the larger we could factor out the x and expand in powers of c/x as our small parameter around $c = 0$. In that case we'd get:

$$f(x) = x^n \left(1 + n\frac{c}{x} + \frac{1}{2!}n(n-1)\left(\frac{c}{x}\right)^2 + \frac{1}{3!}n(n-1)(n-2)\left(\frac{c}{x}\right)^3 + \dots\right) \quad (65)$$

Remember, n is *arbitrary* in this expression but you should also verify that if n is any positive integer, the series terminates and you recover $(c + x)^n$ *exactly*. In this case the "small" requirement is no longer necessary.

We summarize both of these forms of the expansion by the part in the brackets. Let $y < 1$ and n be an arbitrary real or complex number (although in this class we will use only n real). Then:

$$(1 + y)^n = 1 + ny + \frac{1}{2!}n(n-1)y^2 + \frac{1}{3!}n(n-1)(n-2)y^3 + \dots \quad (66)$$

This is the binomial expansion, and is *very* useful in physics.

Quadratics and Polynomial Roots

As noted above, the purpose of using algebra in physics is so that we can take known expressions that e.g. describe laws of nature and a particular problem and transform these "truths" into a "true" statement of the answer by isolating the *symbol* for that answer on one side of an equation.

For linear problems that is usually either straightforward or impossible. For "simple" linear problems (a single linear equation) it is always possible and usually easy. For sets of simultaneous linear equations in a small number of variables (like the ones represented in the course) one can "always" use a mix of composition (substitution) and elimination to find the answer desired[20].

What about solving polynomials of higher degree to find values of their variables that represent answers to physics (or other) questions? In general one tries to arrange the polynomial into a *standard form* like the one above, and then finds the *roots* of the

[20]This is not true in the general case, however. One can, and should, if you are contemplating a physics major, take an *entire college level course* in the methodology of linear algebra in multidimensional systems.

polynomial. How easy or difficult this may be depends on many things. In the case of a *quadratic* (second degree polynomial involving at most the square) one can – and we will, below – derive an *algebraic expression* for the roots of an *arbitrary* quadratic.

For third and higher degrees, our ability to solve for the roots is not trivially general. Sometimes we will be able to "see" how to go about it. Other times we won't. There exist computational methodologies that work for most relatively low degree polynomials but for very high degree general polynomials the problem of factorization (finding the roots) is *hard*. We will therefore work through quadratic forms in detail below and then make a couple of observations that will help us factor a few e.g. cubic or quartic polynomials should we encounter ones with one of the "easy" forms.

In physics, quadratic forms are quite common. Motion in one dimension with constant acceleration (for example) quite often requires the solution of a quadratic in time. For the purposes of deriving the quadratic formula, we begin with the "standard form" of a quadratic equation:

$$ax^2 + bx + c = 0 \tag{67}$$

(where you should note well that $c = a_0$, $b = a_1$, $a = a_2$ in the general polynomial formula given above).

We wish to find the (two) values of x such that this equation is true, given a, b, c. To do so we must rearrange this equation and *complete the square*.

$$
\begin{aligned}
ax^2 + bx + c &= 0 \\
ax^2 + bx &= -c \\
x^2 + \frac{b}{a}x &= -\frac{c}{a} \\
x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} &= \frac{b^2}{4a^2} - \frac{c}{a} \\
(x + \frac{b}{2a})^2 &= \frac{b^2}{4a^2} - \frac{c}{a} \\
(x + \frac{b}{2a}) &= \pm\sqrt{\frac{b^2}{4a^2} - \frac{c}{a}} \\
x &= -\frac{b}{2a} \pm \sqrt{\frac{b^2}{4a^2} - \frac{c}{a}} \\
x_\pm &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad (68)
\end{aligned}
$$

This last result is the well-known *quadratic formula* and its general solutions are *complex numbers* (because the argument of the square root can easily be negative if $4ac > b^2$). In some cases the complex solution is *desired* as it leads one to e.g. a *complex exponential* solution and hence a trigonometric oscillatory function as we shall see in the next section. In other cases we insist on the solution being real, because if it isn't there is no real solution to the problem posed! Experience solving problems of both types is needed so that a student can learn to recognize both situations and use complex numbers to their advantage.

Before we move on, let us note two cases where we can "easily" solve cubic or quartic polynomials (or higher order polynomials) for their roots algebraically. One is when we take the quadratic formula and multiply it by any power of x, so that it can be *factored*, e.g.

$$
\begin{aligned}
ax^3 + bx^2 + cx &= 0 \\
(ax^2 + bx + c)x &= 0 \qquad (69)
\end{aligned}
$$

This equation clearly has the two quadratic roots given above plus one (or more, if the power of x is higher) root $x = 0$. In some cases one can factor a solvable term of the form $(x + d)$ by inspection, but this is generally not easy if it is possible at all without solving for the roots some other way first.

The other "tricky" case follows from the observation that:

$$x^2 - a^2 = (x + a)(x - a) \qquad (70)$$

so that the two roots $x = \pm a$ are solutions. We can generalize this and solve e.g.

$$x^4 - a^4 = (x^2 - a^2)(x^2 + a^2) = (x - a)(x + a)(x - ia)(x + ia) \qquad (71)$$

and find the *four* roots $x = \pm a, \pm ia$. One can imagine doing this for still higher powers on occasion.

In *this* course we will almost never have a problem that cannot be solved using "just" the quadratic formula, perhaps augmented by one or the other of these two tricks, although naturally a diligent and motivated student contemplating a math or physics major will prepare for the more difficult future by reviewing the various factorization tricks for "fortunate" integer coefficient polynomials, such as *synthetic division*. However, such a student should *also* be aware that the general problem of finding all the roots of a polynomial of arbitrary degree is *difficult*[21] . So difficult, in fact, that it is known that no *simple* solution involving only arithmetical operations and square roots exists for degree 5 or greater. However it is generally fairly easy to factor arbitrary polynomials to a high degree of accuracy *numerically* using well-known algorithms and a computer.

Now that we understand both inverse functions and Taylor series expansions and quadratics and roots, let us return to the question asked earlier. What happens if we extend the domain of an inverse function outside of the range of the original function? In general we find that the inverse function has no *real* solutions. Or, we can find as noted above when factoring polynomials that like as not there are no real solutions. But that does not mean that solutions do not exist!

[21]Wikipedia: http://www.wikipedia.org/wiki/Polynomial.

Complex Numbers and Harmonic Trigonometric Functions

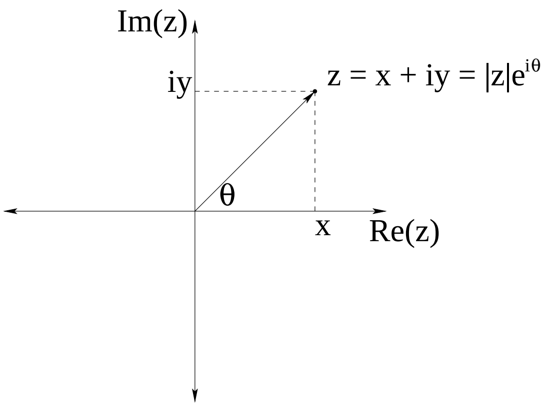


Figure 1: A complex number maps perfectly into the two-dimensional xy coordinate system in both Cartesian and Plane Polar coordinates. The latter are especially useful, as they lead to the Euler representation of complex numbers and complex exponentials.

We already reviewed very briefly the definition of the unit imaginary number $i = +\sqrt{-1}$. This definition, plus the usual rules for algebra, is enough for us to define both the *imaginary numbers* and a new kind of number called a *complex* number z that is the sum of *real and imaginary parts*, $z = x + iy$.

If we plot the real part of z (x) on the one axis and the imaginary part (y) on another, we note that the complex numbers map into a *plane* that looks *just like* the x-y plane in ordinary plane geometry. Every complex number can be represented as an ordered pair of real numbers, one real and one the magnitude of the imaginary. A picture of this is drawn above.

From this picture and our knowledge of the *definitions* of the trigonometric functions we can quickly and easily deduce some *extremely useful and important* True Facts about:

Complex Numbers

This is a very terse review of their most important properties. From the figure above, we can see that an arbitrary complex number z can *always* be written as:

$$
\begin{aligned}
z &= x + iy && (72)\\
&= |z|\left(\cos(\theta) + i|z|\sin(\theta)\right) && (73)\\
&= |z|e^{i\theta} && (74)
\end{aligned}
$$

where $x = |z|\cos(\theta)$, $y = |z|\sin(\theta)$, and $|z| = \sqrt{x^2 + y^2}$. *All complex numbers* can be written as a real amplitude $|z|$ times a complex exponential form involving a phase angle. Again, it is difficult to convey how incredibly useful this result is without devoting an entire book to this alone but for the moment, at least, I commend it to your attention.

There are a variety of ways of deriving or justifying the exponential form. Let's examine just one. If we differentiate z with respect to θ in the second form (73) above we get:

$$\frac{dz}{d\theta} = |z|\left(-\sin(\theta) + i\cos(\theta)\right) = i|z|\left(\cos(\theta) + i\sin(\theta)\right) = iz \tag{75}$$

This gives us a differential equation that is an *identity* of complex numbers. If we multiply both sides by $d\theta$ and divide both sizes by z and integrate, we get:

$$\ln z = i\theta + \text{constant} \tag{76}$$

If we use the inverse function of the natural log (exponentiation of both sides of the equation:

$$
\begin{aligned}
e^{\ln z} &= e^{(i\theta + \text{constant})} = e^{\text{constant}} e^{i\theta}\\
z &= |z|e^{i\theta} && (77)
\end{aligned}
$$

where $|z|$ is basically a constant of integration that is set to be the *magnitude* of the complex number (or its *modulus*) where the complex exponential piece determines its *complex phase.*

There are a number of really interesting properties that follow from the exponential form. For example, consider multiplying two complex numbers a and b:

$$
\begin{aligned}
a &= |a|e^{i\theta_a} = |a|\cos(\theta_a) + i|a|\sin(\theta_a) && (78)\\
b &= |b|e^{i\theta_b} = |b|\cos(\theta_b) + i|b|\sin(\theta_b) && (79)\\
ab &= |a||b|e^{i(\theta_a + \theta_b)} && (80)
\end{aligned}
$$

and we see that multiplying two complex numbers multiplies their *amplitudes* and *adds* their phase angles. Complex multiplication thus *rotates and rescales* numbers in the complex plane.

Trigonometric and Exponential Relations

$$
\begin{aligned}
e^{\pm i\theta} &= \cos(\theta) \pm i\sin(\theta) && (81)\\
\cos(\theta) &= \frac{1}{2}\left(e^{+i\theta} + e^{-i\theta}\right) && (82)\\
\sin(\theta) &= \frac{1}{2i}\left(e^{+i\theta} - e^{-i\theta}\right) && (83)
\end{aligned}
$$

From these relations and the properties of exponential multiplication you can painlessly prove all sorts of trigonometric identities that were immensely painful to prove back in high school

There are a few other trig relations (out of a *lot*[22] that can be derived) that are very useful in physics. For example:

$$\sin(A) \pm \sin(B) = 2\sin\left(\frac{A \pm B}{2}\right)\cos\left(\frac{A \mp B}{2}\right) \tag{84}$$

$$\cos(A) + \cos(B) = 2\cos\left(\frac{A + B}{2}\right)\cos\left(\frac{A - B}{2}\right) \tag{85}$$

$$\cos(A) - \cos(B) = -2\sin\left(\frac{A + B}{2}\right)\sin\left(\frac{A - B}{2}\right) \tag{86}$$

These expressions are used (along with the superposition principle) to add waves with identical amplitudes in the study of beats, interference, and diffraction. However, they are somewhat limited – we don't have a good trigonometric identity for adding *three* or *four* or N sine wave, even when the amplitudes are the same and they angles differ by simple multiples of a fixed phase. For problems like this, we will use *phasor diagrams* to *graphically* find new identities that solve these problems, often with an elegant connection back to complex exponentials.

Power Series Expansions

These can easily be evaluated using the Taylor series discussed in the last section, expanded around the origin $z = 0$, and are an alternative way of seeing that $z = e^{i\theta}$. In the case of exponential and trig functions, the expansions converge for all z, not just small ones (although they of course converge *faster* for small ones).

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \tag{87}$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \dots \tag{88}$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots \tag{89}$$

Depending on where you start, these can be used to prove the relations above. They are most useful for getting expansions for small values of their parameters. For small x (to leading order):

$$e^x \approx 1 + x \tag{90}$$

$$\cos(x) \approx 1 - \frac{x^2}{2!} \tag{91}$$

$$\sin(x) \approx x \tag{92}$$

$$\tan(x) \approx x \tag{93}$$

[22]Wikipedia: http://www.wikipedia.org/wiki/List of trigonometric identities.

We will use these fairly often in this course, so learn them.

An Important Relation

A relation I will state without proof that is very important to this course is that the real part of the $x(t)$ derived above:

$$\Re(x(t)) = \Re(x_{0+}e^{+i\omega t} + x_{0-}e^{-i\omega t}) \tag{94}$$

$$= X_0 \cos(\omega t + \phi) \tag{95}$$

where ϕ is an arbitrary phase. You can prove this in a few minutes or relaxing, enjoyable algebra from the relations outlined above – remember that x_{0+} and x_{0-} are arbitrary *complex* numbers and so can be written in complex exponential form!

Calculus

In this section we present a lightning fast review of calculus. It is *most* of what you need to do well in this course.

Differential Calculus

The slope of a line is defined to be the rise divided by the run. For a *curved* line, however, the slope has to be defined *at a point.* Lines (curved or straight, but not infinitely steep) can always be thought of as *functions* of a single variable. We call the slope of a line evaluated at any given point its *derivative*, and call the process of finding that slope *taking the derivative of the function.*

Later we'll say a few words about multivariate (vector) differential calculus, but that is mostly beyond the scope of this course.

The definition of the derivative of a function is:

$$\frac{df}{dx} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \tag{96}$$

This is the *slope* of the function at the point x.

First, note that:

$$\frac{d(af)}{dx} = a\frac{df}{dx} \tag{97}$$

for any constant a. The constant simply factors out of the definition above.

Second, differentiation is *linear.* That is:

$$\frac{d}{dx}(f(x) + g(x)) = \frac{df(x)}{dx} + \frac{dg(x)}{dx} \tag{98}$$

Third, suppose that $f = gh$ (the product of two functions). Then

$$\begin{aligned}
\frac{df}{dx} &= \frac{d(gh)}{dx} = \lim_{\Delta x \to 0} \frac{g(x+\Delta x)h(x+\Delta x) - g(x)h(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{\Big(g(x) + \frac{dg}{dx}\Delta x)(h(x) + \frac{dh}{dx}\Delta x) - g(x)h(x)\Big)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{\Big(g(x)\frac{dh}{dx}\Delta x + \frac{dg}{dx}h(x)\Delta x + \frac{dg}{dx}\frac{dh}{dx}(\Delta x)^2)\Big)}{\Delta x} \\
&= g(x)\frac{dh}{dx} + \frac{dg}{dx}h(x) \qquad (99)
\end{aligned}$$

where we used the definition above twice and multiplied everything out. If we multiply this rule by dx we obtain the following rule for the differential of a product:

$$d(gh) = g\ dh + h\ dg \qquad (100)$$

This is a *very important result* and leads us shortly to integration by parts and later in physics to things like Green's theorem in vector calculus.

We can easily and directly compute the derivative of a mononomial:

$$\begin{aligned}
\frac{dx^n}{dx} &= \lim_{\Delta x \to 0} \frac{x^n + nx^{n-1}\Delta x + n(n-1)x^{n-2}(\Delta x)^2 \ldots + (\Delta x)^n) - x^2}{\Delta x} \\
&= \lim_{\Delta x \to 0} \left(nx^{n-1} + n(n-1)x^{n-2}(\Delta x) \ldots + (\Delta x)^{n-1}\right) \\
&= nx^{n-1} \qquad (101)
\end{aligned}$$

or we can derive this result by noting that $\frac{dx}{dx} = 1$, the product rule above, and using induction. If one assumes $\frac{dx^n}{dx} = nx^{n-1}$, then

$$\begin{aligned}
\frac{dx^{n+1}}{dx} &= \frac{d(x^n \cdot x)}{dx} \\
&= nx^{n-1} \cdot x + x^n \cdot 1 \\
&= nx^n + x^n = (n+1)x^n \qquad (102)
\end{aligned}$$

and we're done.

Again it is beyond the scope of this short review to *completely* rederive all of the results of a calculus class, but from what has been presented already one can see how one can systematically proceed. We conclude, therefore, with a simple table of useful derivatives and results in summary (including those above):

$$
\begin{aligned}
\frac{da}{dx} &= 0 \qquad a \text{ constant} && (103)\\
\frac{d(af(x))}{dx} &= a\frac{df(x)}{dx} \qquad a \text{ constant} && (104)\\
\frac{dx^n}{dx} &= nx^{n-1} && (105)\\
\frac{d}{dx}(f(x)+g(x)) &= \frac{df(x)}{dx} + \frac{dg(x)}{dx} && (106)\\
\frac{df}{dx} &= \frac{df}{du}\frac{du}{dx} \qquad \text{chain rule} && (107)\\
\frac{d(gh)}{dx} &= g\frac{dh}{dx} + \frac{dg}{dx}h \qquad \text{product rule} && (108)\\
\frac{d(g/h)}{dx} &= \frac{\frac{dg}{dx}h - g\frac{dh}{dx}}{h^2} && (109)\\
\frac{de^x}{dx} &= e^x && (110)\\
\frac{de^{(ax)}}{dx} &= ae^{ax} \qquad \text{from chain rule, } u = ax && (111)\\
\frac{d\sin(ax)}{dx} &= a\cos(ax) && (112)\\
\frac{d\cos(ax)}{dx} &= -a\sin(ax) && (113)\\
\frac{d\tan(ax)}{dx} &= \frac{a}{\cos^2(ax)} = a\sec^2(ax) && (114)\\
\frac{d\cot(ax)}{dx} &= -\frac{a}{\sin^2(ax)} = -a\csc^2(ax) && (115)\\
\frac{d\ln(x)}{dx} &= \frac{1}{x} && (116)\\
& && (117)
\end{aligned}
$$

There are a few more integration rules that can be useful in this course, but nearly all of them can be derived in place using these rules, especially the chain rule and product rule.

Integral Calculus

With differentiation under our belt, we need only a few definitions and we'll get integral calculus for free. That's because integration is *antidifferentiation*, the *inverse process* to differentiation. As we'll see, the derivative of a function is unique but its integral has *one free choice* that must be made. We'll also see that the (definite) integral of a function in one dimension is the *area underneath the curve.*

There are lots of ways to facilitate derivations of integral calculus. Most calculus books begin (appropriately) by drawing pictures of curves and showing that the area beneath them can be evaluated by summing small discrete sections and that by means of a limiting process that area is equivalent to the integral of the functional curve. That is, if $f(x)$ is some curve and we wish to find the area beneath a segment of it (from $x = x_1$ to $x = x_2$ for example), one small piece of that area can be written:

$$\Delta A = f(x)\Delta x \tag{118}$$

The total area can then be approximately evaluated by piecewise summing N rectangular strips of width $\Delta x = (x_2 - x_1)/N$:

$$A \approx \sum_{n=1}^{N} f(x_1 + n \cdot \Delta x)\Delta x \tag{119}$$

(Note that one can get slightly different results if one centers the rectangles or begins them on the low side, but we don't care.)

In the limit that $N \to \infty$ and $\Delta x \to 0$, two things happen. First we note that:

$$f(x) = \frac{dA}{dx} \tag{120}$$

by the definition of derivative from the previous section. The function $f(x)$ is the formal derivative of the function representing the area beneath it (independent of the limits as long as x is in the domain of the function.) The second is that we'll get tired adding teensy-weensy rectangles in infinite numbers. We therefore make up a *special symbol* for this infinite limit sum. Σ clearly stands for sum, so we change to another stylized "ess", $\int$, to *also* stand for sum, but now a continuous and infinite sum of all the infinitesimal pieces of area within the range. We now write:

$$A = \int_{x_1}^{x_2} f(x)dx \tag{121}$$

as an *exact* result in this limit.

The beauty of this simple approach is that we now can do the following algebra, over and over again, to formulate integrals (sums) of some quantity.

$$\begin{aligned} \frac{dA}{dx} &= f(x) \\ dA &= f(x)dx \\ \int dA &= \int f(x)dx \\ A &= \int_{x_1}^{x_2} f(x)dx \end{aligned} \tag{122}$$

This areal integral is called a *definite integral* because it has definite upper and lower bounds. However, we can *also* do the integral with a *variable* upper bound:

$$A(x) = \int_{x_0}^{x} f(x')dx' \tag{123}$$

where we indicate how A *varies* as we change x, its upper bound.

We now make a clever observation. $f(x)$ is clearly the function that we get by differentiating this integrated area with a fixed lower bound (which is still arbitrary) with respect to the variable in its upper bound. That is

$$f(x) = \frac{dA(x)}{dx} \tag{124}$$

This slope must be the *same* for all possible values of x_0 or this relation would not be correct and unique! We therefore conclude that all the various functions $A(x)$ that can stand for the area differ *only by a constant* (called the constant of integration):

$$A'(x) = A(x) + C \tag{125}$$

so that

$$f(x) = \frac{dA'(x)}{dx} = \frac{dA(x}{dx} + \frac{dC}{dx} = \frac{dA(x)}{dx} \tag{126}$$

From this we can conclude that the *indefinite* integral of $f(x)$ can be written:

$$A(x) = \int^x f(x)dx + A_0 \tag{127}$$

where A_0 is the constant of integration. In physics problems the constant of integration must usually be evaluated algebraically from information given in the problem, such as initial conditions.

From this simple definition, we can transform our *existing* table of derivatives into a table of (indefinite) integrals. Let us compute the integral of x^n as an example. We wish to find:

$$g(x) = \int x^n dx \tag{128}$$

where we will ignore the constant of integration as being irrelevant to this process (we can and should always add it to one side or the other of any formal indefinite integral unless we can see that it is zero). If we differentiate both sides, the differential and integral are inverse operations and we know:

$$\frac{dg(x)}{dx} = x^n \tag{129}$$

Looking on our table of derivatives, we see that:

$$\frac{dx^{n+1}}{dx} = (n+1)x^n \tag{130}$$

or

$$\frac{dg(x)}{dx} = x^n = \frac{1}{n+1}\frac{dx^{n+1}}{dx} \tag{131}$$

and hence:

$$g(x) = \int^x x^n dx = \frac{1}{n+1}x^{n+1} \tag{132}$$

by inspection.

Similarly we can match up the other rules with integral equivalents.

$$\frac{d(af(x))}{dx} = a\frac{df(x)}{dx} \tag{133}$$

leads to:

$$\int af(x)dx = a\int f(x)dx \tag{134}$$

A very important rule follows from the rule for differentiating a product. If we integrate both sides this becomes:

$$\int d(gh) = gh = \int gdh + \int hdg \tag{135}$$

which we often rearrange as:

$$\int gdh = \int d(gh) - \int hdg = gh - \int hdg \tag{136}$$

the rule for *integration by parts* which permits us to throw a derivative from one term to another in an integral we are trying to do. This turns out to be very, very useful in evaluating many otherwise extremely difficult integrals.

If we assemble the complete list of (indefinite) integrals that correspond to our list of derivatives, we get something like:

$$\begin{aligned}
\int 0\ dx &= 0 + c = c \qquad \text{with } c \text{ constant} && (137)\\
\int af(x)dx &= a\int f(x)dx && (138)\\
\int x^n dx &= \frac{1}{n+1}x^{n+1} + c && (139)\\
\int (f+g)dx &= \int f\ dx + \int g\ dx && (140)\\
\int f(x)dx &= \int f(u)\frac{dx}{du}du \qquad \text{change variables} && (141)\\
\int d(gh) &= gh = \int gdh + \int hdg \qquad \text{or} && (142)\\
\int gdh &= gh - \int hdg \qquad \text{integration by parts} && (143)\\
\int e^x dx &= e^x + a \qquad \text{or change variables to} && (144)\\
\int e^{ax} dx &= \frac{1}{a}\int e^{ax} d(ax) = \frac{1}{a}e^{ax} + c && (145)\\
\int \sin(ax)dx &= \frac{1}{a}\int \sin(ax)d(ax) = \frac{1}{a}\cos(ax) + c && (146)\\
\int \cos(ax)dx &= \frac{1}{a}\int \cos(ax)d(ax) = -\frac{1}{a}\sin(ax) + c && (147)\\
\int \frac{dx}{x} &= \ln(x) + c && (148)\\
& && (149)
\end{aligned}$$

It's worth doing a couple of examples to show how to do integrals using these rules. One integral that appears in many physics problems in E&M is:

$$\int_0^R \frac{r\ dr}{(z^2+r^2)^{3/2}} \tag{150}$$

This integral is done using *u substitution* – the chain rule used backwards. We look at it for a second or two and note that if we let

$$u = (z^2+r^2) \tag{151}$$

then

$$du = 2r dr \tag{152}$$

and we can rewrite this integral as:

$$\begin{aligned} \int_0^R \frac{r\ dr}{(z^2+r^2)^{3/2}} &= \frac{1}{2}\int_0^R \frac{2r\ dr}{(z^2+r^2)^{3/2}} \\ &= \frac{1}{2}\int_{z^2}^{(z^2+R^2)} u^{-3/2}\ du \\ &= -u^{-1/2}\Big|_{z^2}^{(z^2+R^2)} \\ &= \frac{1}{z} - \frac{1}{(z^2+R^2)^{1/2}} \end{aligned} \tag{153}$$

The lesson is that we can often do complicated looking integrals by making a suitable u-substitution that reduces them to a simple integral we know off of our table.

The next one illustrates both integration by parts and doing integrals with infinite upper bounds. Let us evaluate:

$$\int_0^\infty x^2 e^{-ax} dx \tag{154}$$

Here we identify two pieces. Let:

$$h(x) = x^2 \tag{155}$$

and

$$d(g(x)) = e^{-ax}dx = -\frac{1}{a}e^{-ax}d(-ax) = -\frac{1}{a}d(e^{-ax}) \tag{156}$$

or $g(x) = -(1/a)e^{-ax}$. Then our rule for integration by parts becomes:

$$\begin{aligned} \int_0^\infty x^2 e^{-ax}dx &= \int_0^\infty h(x)dg \\ &= h(x)g(x)\Big|_0^\infty - \int_0^\infty g(x)dh \\ &= -\frac{1}{a}x^2e^{-ax}\Big|_0^\infty + \frac{1}{a}\int_0^\infty e^{-ax}2xdx \\ &= \frac{2}{a}\int_0^\infty xe^{-ax}dx \end{aligned} \tag{157}$$

We repeat this process with $h(x) = x$ and with $g(x)$ unchanged:

$$\begin{aligned}
\int_0^\infty x^2 e^{-ax} dx &= \frac{2}{a}\int_0^\infty x e^{-ax} dx \\
&= -\frac{2}{a^2} x e^{-ax}\Big|_0^\infty + \frac{2}{a^2}\int_0^\infty e^{-ax} dx \\
&= \frac{2}{a^2}\int_0^\infty e^{-ax} dx \\
&= -\frac{2}{a^3}\int_0^\infty e^{-ax} d(-ax) \\
&= -\frac{2}{a^3} e^{-ax}\Big|_0^\infty = \frac{2}{a^3}
\end{aligned} \tag{158}$$

If we work a little more generally, we can show that:

$$\int_0^\infty x^n e^{-ax} dx = \frac{(n+1)!}{a^n} \tag{159}$$

This is just one illustration of the power of integration by parts to help us do integrals that on the surface appear to be quite difficult.

Vector Calculus

This book will not use a great deal of vector or multivariate calculus, but a *little* general familiarity with it will greatly help the student with e.g. multiple integrals or the idea of the force being the negative gradient of the potential energy. We will content ourselves with a few definitions and examples.

The first definition is that of the *partial derivative*. Given a function of many variables $f(x, y, z...)$, the partial derivative of the function with respect to (say) x is written:

$$\frac{\partial f}{\partial x} \tag{160}$$

and is just the regular derivative of the *variable form* of f as a function of all its coordinates with respect to the x coordinate only, holding all the *other* variables constant even if they are not independent and vary in some known way with respect to x.

In many problems, the variables *are* independent and the partial derivative is equal to the regular derivative:

$$\frac{df}{dx} = \frac{\partial f}{\partial x} \tag{161}$$

In other problems, the variable y might *depend* on the variable x. So might z. In that case we can form the *total* derivative of f with respect to x by including the variation of f caused by the variation of the other variables as well (basically using the chain rule and composition):

$$\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} + ... \tag{162}$$

Note the different *full* derivative symbol on the left. This is called the "total derivative" with respect to x. Note also that the independent form follows from this second form because $\frac{\partial y}{\partial x} = 0$ and so on are the *algebraic* way of saying that the coordinates are independent.

There are several ways to form vector derivatives of functions, especially *vector* functions. We begin by defining the *gradient* operator, the basic vector differential form:

$$\vec{\boldsymbol{\nabla}} = \frac{\partial}{\partial x}\hat{\boldsymbol{x}} + \frac{\partial}{\partial y}\hat{\boldsymbol{y}} + \frac{\partial}{\partial z}\hat{\boldsymbol{z}} \tag{163}$$

This operator can be applied to a scalar multivariate function f to form its gradient:

$$\vec{\boldsymbol{\nabla}} f = \frac{\partial f}{\partial x}\hat{\boldsymbol{x}} + \frac{\partial f}{\partial y}\hat{\boldsymbol{y}} + \frac{\partial f}{\partial z}\hat{\boldsymbol{z}} \tag{164}$$

The gradient of a function has a magnitude equal to its *maximum* slope at the point in any possible direction, pointing in the direction in which that slope is maximal. It is the "uphill slope" of a curved surface, basically – the word "gradient" *means* slope. In physics this directed slope is *very* useful.

If we wish to take the vector derivative of a vector function there are two common ways to go about it. Suppose $\vec{\boldsymbol{E}}$ is a vector function of the spatial coordinates. We can form its *divergence*:

$$\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{E}} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \tag{165}$$

or its *curl*:

$$\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{E}} = (\frac{\partial E_y}{\partial z} - \frac{\partial E_z}{\partial y})\hat{\boldsymbol{x}} + (\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z})\hat{\boldsymbol{y}} + (\frac{\partial E_x}{\partial y} - \frac{\partial E_y}{\partial x})\hat{\boldsymbol{z}} \tag{166}$$

These operations are extremely important in physics courses, especially the more advanced study of electromagnetics, where they are part of the differential formulation of Maxwell's equations, but we will not use them in a required way in this course. We'll introduce and discuss them and work a rare problem or two, just enough to get the *flavor* of what they mean onboard to front-load a more detailed study later (for majors and possibly engineers or other advanced students only).

Multiple Integrals

The last bit of multivariate calculus we need to address is integration over *multiple* dimensions. We will have many occasions in this text to integrate over *lines*, over *surfaces*, and over *volumes* of space in order to obtain quantities. The integrals themselves are not difficult – in this course they can *always* be done as a series of one, two or three ordinary, independent integrals over each coordinate one at a time with the others held "fixed". This is not always possible and multiple integration can get much more difficult, but we *deliberately* choose problems that illustrate the general idea of integrating over a volume while still remaining accessible to a student with fairly modest calculus skills, no more than is required and reviewed in the sections above.

[Note: This section is not yet finished, but there are examples of all of these in context in the relevant sections below. Check back for later revisions of the book PDF (possibly after contacting the author) if you would like this section to be filled in urgently.]