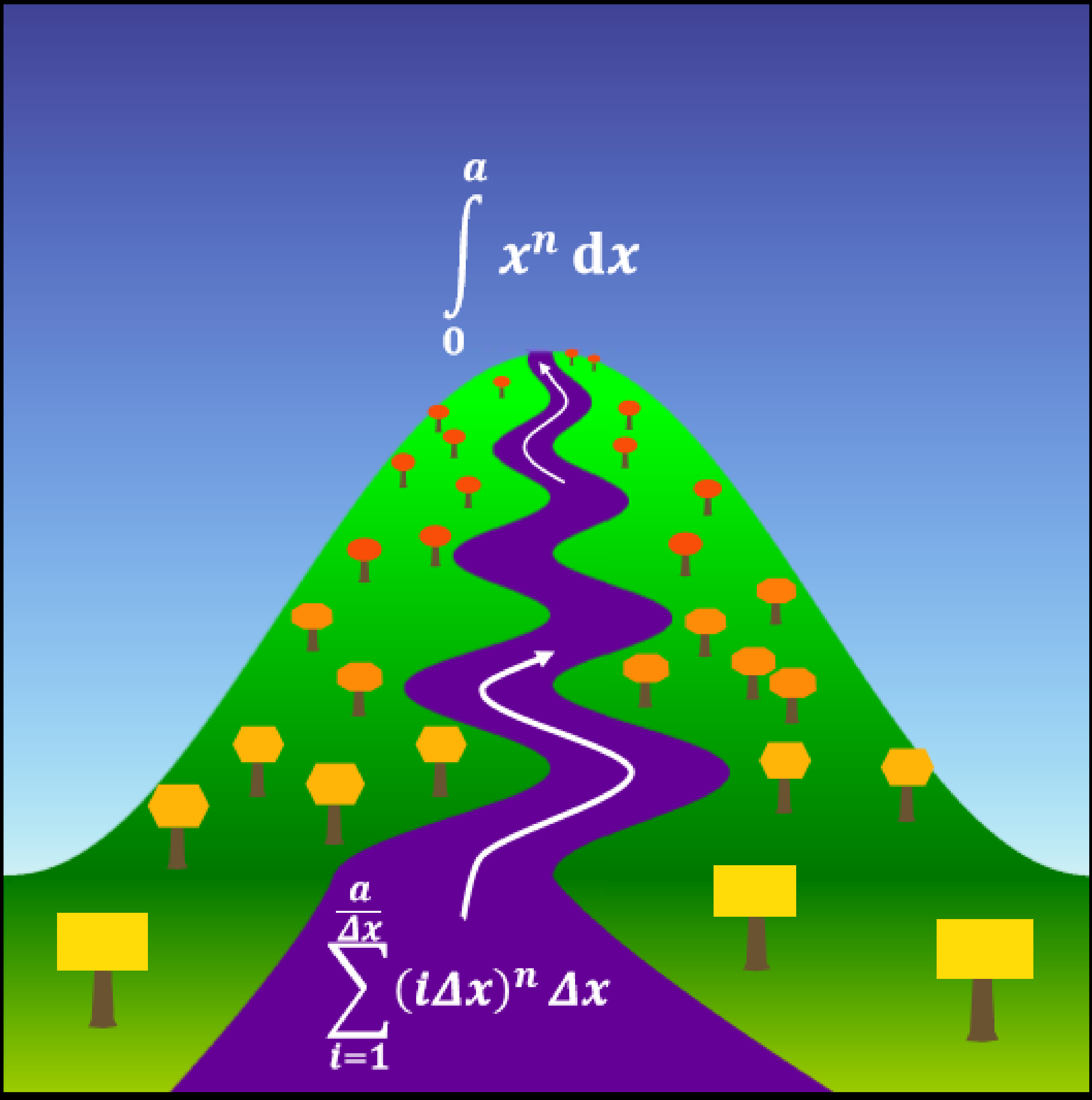
CALCULUS

YOUR ROYAL ROAD TO GENIUS



STANLEY DAVID GEDZELMAN

Calculus: Your Royal Road to Genius

Stanley David Gedzelman

Professor Emeritus of Atmospheric Sciences
City College of New York

Stanley David Gedzelman Books

Revised 09 Sept 2020

ISBN-10:0991197607
ISBN-13:978-0-9911976-0-6

QA 303.G___ 2014
515 GED

To my Granddaughters,
Alexia Ruby, Naomi Savannah, and Charlotte Aviva
(and any that may follow).
So that as you conquer your world you may get to know
Calculus (and your Grandpa) better.

And to all of you with the courage to tackle Calculus
and help make the world better!

PREFACE

You who are about to embark on the perilous journey to and through Calculus, I salute you. I have written this book for you. According to myth, Ptolemy I asked Euclid for a shortcut to learn Geometry. Euclid replied, "Sire, there is no Royal Road to Geometry," referring both to the difficulty of the subject and to the Royal Road that facilitated travel between Sardis and Susa. So, be prepared for hard, long work to absorb the difficult concepts involved in mastering Calculus.

But if you master Calculus, one of humankind's great intellectual adventures and triumphs, you will have transformed yourself into a genius who can learn and do almost anything. Calculus will have been your Royal Road to Genius. So, join me on that road.

This book contains almost everything covered in a two or three semester Calculus course. I started writing it 21 years ago and put it down soon thereafter as other projects took precedence. But when I took up Substitute Teaching in the Fall of 2012 (after a long career as a Professor of Atmospheric Sciences) I was shocked at the excessive length of math textbooks. I couldn't see the forest for the (un)focus boxes. That inspired me to resurrect my dusty old manuscript, complete it, and keep it short. Unlike so many authors I didn't have to satisfy multiple school districts. I only had to satisfy myself and you. Besides, I find that writing a book is a good way to review and even learn a subject.

Well, that's just about all I have to say now. This preface is short and so is the book. I had fun and I hope you will. But I repeat, it will take work to become the genius you are!

Stanley David Gedzelman
02 January 2014

Oh, one more thing! Read the next page for a head start Overview of Calculus.
09 September 2020

Overview: What is Calculus?

What is Calculus and what does it do?

Calculus is the mathematics of continuous change. We use it to calculate,

1: *Rates of change*, called Derivatives in Calculus, such as how fast a rocket is moving, and how fast it *will be* moving.

2: *Totals,* called Integrals in Calculus, such as how far and to where the rocket *will* travel in a given time period.

This means that **Calculus might be called the math of making predictions**. Calculus can also be used backwards to infer the past.

Calculus might also be called the math of curves. Geometry and Trigonometry (which are prerequisites for Calculus) describe properties of many curves such as circles, ellipses, parabolas, hyperbolas, and sine and cosine waves. But Geometry and Trigonometry cannot *prove* many of their properties – only Calculus can. For example, Geometry and Trigonometry must *assume* at least three of the circle's fundamental properties, namely,

1: Its area, $\boldsymbol{A} = \pi \boldsymbol{r}^2$.

2: Its circumference, $\boldsymbol{C} = 2\pi \boldsymbol{r}$.

3: The value of pi, $\pi \approx 3.14159265.....$

Calculus is the math allows us to calculate and prove these values.

How does Calculus do these things? It does them by taking **infinitesimal** steps and, if needed, adding them up.

Note that both words, *Calculus* and *calculate*, derive from the Latin for small pebble, which is how people calculated (as with an abacus) in the old days.

Now, I freely admit that none of the above noble words about what Calculus can do will help *you do* anything in Calculus. All that those noble words do is to give you some notion, however **infinitesimal** (pun intended), of what you will be learning and doing as you study and perhaps struggle through Calculus.

So! All you have to do to learn Calculus is to read the rest of this book!

Stanley David Gedzelman
25 August 2020

CALCULUS: YOUR ROYAL ROAD TO GENIUS
TABLE OF CONTENTS

CALCULUS: YOUR ROYAL ROAD TO GENIUS
CHAPTER 0: BEFORE CALCULUS (BC)

0.1 Introduction
Pattern and its Language: Mathematics

Mathematics is the language of pattern, and pattern appears all around us, as in Figure 0-1. Galileo captured math's essence and importance when he wrote,

Figure 0-1 The logarithmic spiral of the Nautilus shell.

> "The universe cannot be read until we have learned the language and become familiar with the characters in which it is written. It is written in mathematical language, and the letters are triangles, circles and other geometrical figures, without which means it is humanly impossible to comprehend a single word. Without these, one is wandering about in a dark labyrinth."

Math is indeed a language. This has made it so hard to learn and master, especially as it is abstract and as we do not speak it routinely. But its beauty and power have long been acknowledged, and sometimes even revered. (Just think of the magical or diabolical properties often attributed to numbers such as 3, 7, 18, or 666.)

The respect we accord math is reflected in its name. When you look up the origin of the word, mathematics, you might be surprised to find, as I was, that it derives from the Greek, μανθάνω (methaino) - **to learn**. That's pretty darned all encompassing. It is why *polymath* is the word for a person with wide knowledge. It is also why many of us think of people who are outstanding at math as the smartest people of all even though we might call them nerds or other less complimentary names. In the 3rd book of ***Gulliver's Travels***, Jonathan Swift satirized mathematicians as totally divorced from the real world. How ironic it is that while mathematicians may well be absent minded, and math is abstract, it is central to the practical worlds of science, technology and business.

Math's importance in society today is so pervasive it is almost impossible to appreciate fully. It invades, occupies, permeates, and regulates almost every aspect of our lives and of nature, as Galileo noted. And because math is the language of pattern, it is concerned not only with immutable truth but also with beauty.

Why then is math perhaps the most hated, least popular subject? Its immutability is sometimes offered as a reason, but more often, is its difficulty and abstractness. So, like sour grapes, many people accuse math of being irrelevant to their lives, pointing out that they will never again solve or even write an equation the second they have finished their last math course. But we can't escape making measurements and calculating finances or we will fall off the bridge and go bankrupt.

I don't think anyone can make you fall in love with math (or with anything). You must have the potential within yourself. Perhaps someone can point the way to your potential love of math. I will try. How will I go about it? First, I will address one main obstacle – the feeling of inadequacy and stupidity. You should know that we are all in the same boat. So, never feel inferior to the other silent students in your Calculus class (who are at least as terrified as you) or to the students who ask pompous questions designed to impress, or to pretentious professors who may well have struggled for years to learn math and glibly cover up their pasts. The worst thing about these negative feelings is that they reduce your concentration and then become self-fulfilling prophecies. The real truth is not that we are stupid, it is that the subject is difficult and takes time and effort.

One way to partly bypass feelings of stupidity is to show some of math's magnificent applications. Too often, math is presented as an abstract intellectual triumph divorced from its practical sources. Nonsense! Practical was there at the birth of Calculus. Math has always been inspired and forced by down to earth applications such as money, nature (astronomy and the orbits of the planets), architecture, and engineering. Practical is good! If it helps to think of problems in terms of money, think in terms of money. What is wrong with counting with your fingers? Nothing! After all, where do you think base, 10 and the decimal system came from? You got it, our fingers! Well, maybe our toes.

I also try to highlight and demystify the many tricks or steps of genius that seem to come out of nowhere, and show how they were motivated. Take comfort that the hours you spend trying to master these tricks and difficult concepts are nothing compared to the years of sustained, concentrated effort the founders spent to make their discoveries.

I also will present math 1: without the distraction boxes (misnamed focus boxes) that deflect attention and mask rather than illustrate the main point and 2: with only a few problems instead of the enormous lists of repetitive problems that bulk up most texts and can keep you busy for the rest of your lives. As a result, most texts have grown to gargantuan proportions, enough to raise sea level if they were thrown into the ocean, which is where some of them belong. But use them for their solved problems.

A Personal Statement

When I was young I was almost convinced I was a math genius. I thought a genius is someone who learns without working. I learned most of my high school math without working very much or very well when I did work. While the teacher crawled through the high school syllabus I snobbishly sneered at all the struggling and gasping morons who just couldn't catch on, while everything – well, almost everything was trivially obvious to me. I had no idea how arrogant and immature I was.

Then I took Calculus. I used my old method of study – no study where possible and poor study otherwise. The first week Calculus seemed to be a breeze. But I soon began to fall behind. After 3 weeks I was 2 weeks behind. After 6 weeks I was 7 weeks behind. For a year I persisted in my old habits and illusions despite all facts to the contrary. It took me 4 years of nightmares and hard work to catch up in math. Calculus was my first infinitesimal step on my long, not so royal road to realize how childish my definition of genius was. To paraphrase Euclid, The Royal Road to Calculus is...HARD WORK!

Calculus also made me feel desperate. You may be reading this because you feel the same way. Your first test was a disaster and you thought you knew the stuff. Now the teacher is beginning to lecture in what seems to be another language, and it is. So you are trying this book to rescue yourself. Merely acquiring another book won't elevate your math status one iota! To get smart you will have to read <u>actively</u>. Repeat all the derivations and problems and then solve more problems. You absolutely must persist in the struggle because it takes time and effort to rewire our brains, which is what we all have to do to achieve mastery in any subject. And if you master Calculus, you can master anything. Calculus is your Royal Road to Genius.

Of course, books matter. Whenever you feel either humbled or triumphant, you will see that I am with you all the way and am rooting for you. Calculus can be so tough, especially when a bad book is combined with a bad teacher, that you deserve all the help you can get. I loathed my Freshman Calculus text – it was written poorly by the math chairman – and caused me undue pain. I discovered later after much needless pain that there were many far superior Calculus books. I assure you I will feel greatly rewarded if I reduce any pain and increase any pleasure Calculus might cause you. Remember that much of my motivation for writing this book is to spare you some of the pain I suffered and, in a way, to recapture time for myself. I think that if I could have read my own book, I would have learned Calculus sooner, better, and with much less pain.

One last story to give you hope. As a child, Isaac Newton showed some practical genius, making devices such as sundials and windmills, but he didn't know *a bit* of math. The he went to Cambridge University where he began to waste his time with the medieval scholastic curriculum. Possibly to find something of real value he began to read a book on astrology. But a statement about the trigonometry of the Constellations perplexed him. Thus began his study of math. He glanced at geometry and said something like, "This stuff is obvious." But when he tried to understand trigonometry without geometry he couldn't even begin. He must then have said, "Oops, I guess there is much more to geometry than I thought, so I must study it." After geometry, he could tackle trigonometry. Within two years of this humble start he invented Calculus! If Newton could invent Calculus under such circumstances, you have no excuse for not trying to learn it. And, if you feel weak about algebra, geometry and, trigonometry, don't be embarrassed or hesitate to review them. I wrote Chapter 0 to give you that review.

Get Ready: Get Set

Because color stands out Warnings and Alerts in this book are printed in red while important statements of general value are printed in blue, green, or purple. Important terms may be **bold** or underlined. I did all this so that you won't have to highlight or underline, which is a waste of time and unaesthetic. Underlining did not help me learn and made the books ugly. Often when I looked back weeks later I had no idea why I highlighted what I did. If you want to learn a difficult point, write it yourself on a separate sheet of blank paper. Finally, highlighting a printed book lowers its resale value.

Also, Calculus is loaded with symbols that may look strange to you for quite a while but you will see over and over again. It really helps to get used to them. Table 0-1 lists some of these symbols. Some, such as the differential, the derivative, and the integral may not mean anything to you now, but they form the core of Calculus so they will.

Table 0-1 Some Useful Symbols in This Book: Learn Them

Symbol	Meaning	Example
∞	Infinity (endless)	Length of most math texts.
$\Rightarrow$	Implies	$2A = 2B \quad \Rightarrow \quad A = B$
$\rightarrow$	Approaches	As a number, $n \rightarrow \infty$, $1/n \rightarrow 0$
$\equiv$	Exactly equal to by definition	$F \equiv$ Force
$\approx$	Approximately equal	(pi) $\pi \approx 3.1415926$

$\times$ or $\cdot$	* Multiplied by	$3\times4 = 3\cdot4 = 12$
$\propto$	Proportional to	$2\cdot x \propto x$
!	Factorial	$5! = 1\times2\times3\times4\times5 = 120$
x_1	Term with Subscript x_1 is first value of x	Area = $A_{rectangle} = b\times h$ = base$\times$height
Δ	(delta) Difference of or change of	$x_1 = 3$, $x_2 = 3.1$ $\Delta x = x_2 - x_1 = 0.1$
d	Differential = Infinitesimal Difference of	
$\frac{dy}{dx}$	Derivative = Rate of Change, Slope, ratio of differentials	$y = x^3 \Rightarrow \frac{dy}{dx} = 3x^2$
Σ	Sum $\sum_{i=1}^{n} a_i = a_1 + a_2 + a_3 + ... + a_n$	$\sum_{i=1}^{n=5} i = 1+2+3+4+5 = 15$
$\int$	Integral = Area, Volume, Cumulative Total	$\int_4^7 x^2 dx = \frac{7^3 - 4^3}{3}$

* $\times$ and $\cdot$ also represent the cross and dot product (see Section 6-13 on Vector Math).

Did you just skip the Table? Most of us do. Unfortunately most of us forget where the Table is or even that it exists and then we are lost when we need its information. If you remember most or all of these symbols, great! If you don't, here is a brief...

Quiz. (Don't get scared - just use the Table!) Question: How do we write 3 factorial and what is its value? (Answers: 3! and 6 = $3\times2\times1$.) The next question concerns subscripts, which are often attached to variables or constants when there is more than one value. Question: How might we use subscripts (and symbols) to write that Barry is 70 inches tall, Carrie is 58 inches tall and Harry is 63 inches tall? (Possible answers: h_B = 70", h_C = 58" and h_H = 63".) I used the words, "might" and "Possible" because there are alternatives! I chose to use *h* for height and to abbreviate B for Barry, etc.

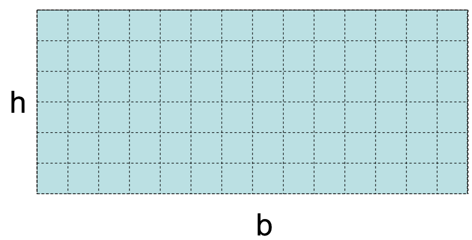
Why We Need Background Mathematics (and Why this is Chapter 0).

I called this Chapter 0 because you already got credit for geometry, algebra and trigonometry. (Analytic Geometry is probably new to you.) But, credit or no, you can't do Calculus until you really know these prerequisites, as Newton found out.

This chapter contains formulas and derivations you will need in Calculus. If you don't understand them or have forgotten them don't worry. You'll see them so often they will eventually become your friends. But, you may say, "Forget the derivations! Just give me the formulas!" That's foolish because derivations often give a general understanding that helps make math stick. Visualizing shapes and patterns and linking them to the equations also helps make math stick. That's a good reason to start with geometry, to draw all graphs in this book, and to seek out relevant animations on the web.

0.2 Areas, Angles and Lengths of Polygons: The Pythagorean Theorem

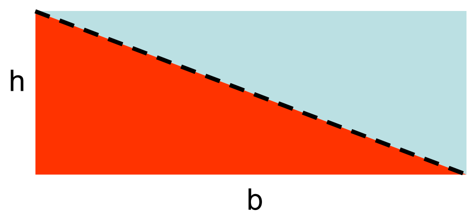
Calculus is often used to find areas of curves such as circles and parabolas. Area formulas for rectangles, triangles and parallelograms are so simple they are given to kids and certainly don't need Calculus. I will start simply, going step-by-step. The area of a rectangle (see Figure 0-2) is length $\times$ width or base (b) $\times$ height (h), (Eqn. 0-1). Of course you can count squares but that would be a waste of time. Multiplying is faster.



$$A_{\text{rectangle}} = bh = A_{prl}$$ Eqn. 0-1

Figure 0-2 A rectangle with base, *b* and height, *h*. To find area, count squares or multiply!

When you draw the diagonal connecting opposite corners of a rectangle (see Figure 0-3), you divide the rectangle in two equal right triangles. So the area of a <u>right</u> triangle is half the area of the rectangle or, ½*bh* (Eqn. 0-2).



$$A_{\text{triangle}} = \frac{1}{2}bh$$ Eqn. 0-2

Figure 0-3 A diagonal line (dashed) divides a rectangle in two right triangles

Figure 0-4 shows that any triangle can be divided into two right triangles by drawing the *altitude*, a line from the corner point or *vertex* of the largest angle at right angles to the opposite side. As a result, Figure 0-4 also shows that the area formula, Eqn. 0-2, is true for all triangles.

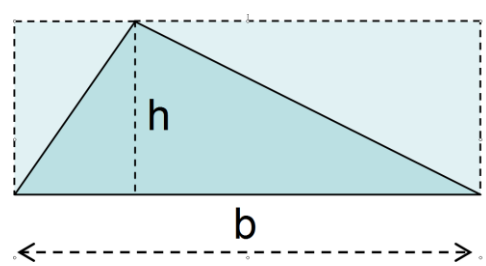


Figure 0-4 Any triangle can be split in two right triangles and rotated so that the base opposite the largest angle is horizontal.

Hint: Rotate the triangle so that the side opposite the largest angle is horizontal and is the base.

Figure 0-5 shows that any parallelogram can be transformed into a rectangle by cutting off a right triangle on one side and patching it to the opposite side. As a result, Figure 0-5 also shows that the area formula for rectangles, Eqn. 0-1, also applies to parallelograms.

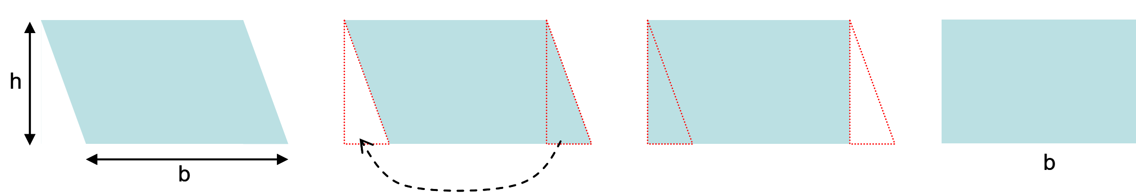


Figure 0-5 Transforming a parallelogram to a rectangle.

Problem: Estimate how many ears of corn will grow on a rectangular field with *b* = 850 ft, and *h* = 600 ft?

Information: Farmers plant corn about ½ ft apart in rows about 3 ft apart. Most stalks yield 1 ear of corn. Assume this is the average.
Solution: The area of the field is

$$A_{\text{rect}} = bh = 850 \times 600 = 510000 \ \text{ft}^2$$

The area one ear needs is, $A_{ear} \approx \frac{1}{2} \times 3 \approx 1.5$ ft^2/ear. This leads to the proportion,

$$\frac{N_{ears}}{A_{rect}} = \frac{1_{ear}}{A_{ear}}$$

Cross multiplying yields the solution,

$$N_{ears} = 1_{ear} \frac{A_{rect}}{A_{ear}} = \frac{510000}{1.5} = 340000$$

Farmer's Note: Since there are ≈ 150 ears of (shelled) corn per bushel, corn yield is ≈ 340,000/ 150 = 2267 bushels. Of course, corn yields decrease in a drought, but that is the farmer's problem, not our math problem (until prices go up for everything).

Angles

An angle is a measure of direction or change of direction. It is almost always linked to the circle. Figure 0-6 shows a circle with two angles, *A* and *B*. Angle, *A* ≈ 172°. Angle, *B* ≈ 42°. Why do we measure angles in degrees? The Egyptians divided the year into 360 days plus a few extra days. The number, 360 stuck as a cycle or a complete turn.

That was the origin of 360 degrees (360°) in a circle. After years of math in school we all get so used to 360° that it seems to have come from God.

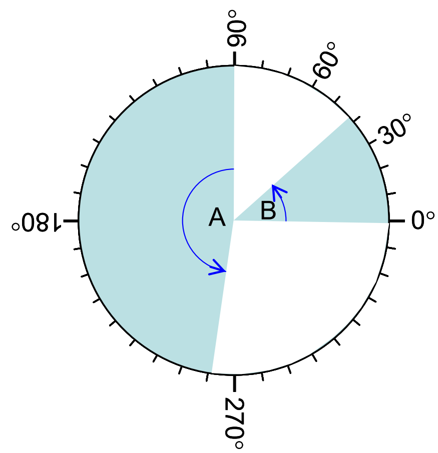


Figure 0-6 A circle with angles marked in degrees.

It doesn't!

Figure 0-7 shows that a line that rotates (around the black dot) so that it faces the opposite direction turns by half a circle or 180° (a *straight angle*) or two *right* (90°) *angles*. It also shows that the angles of a triangle add up to 180°. Because the horizontal dashed and solid lines are parallel, both angles marked *A* are equal as are both angles marked *B*. (Perhaps you remember that they are called *alternate interior angles*.)

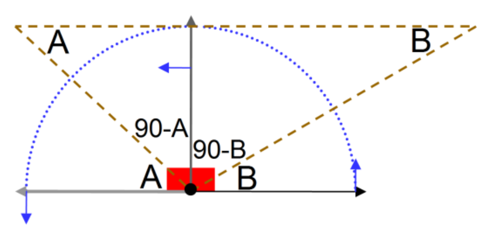


Figure 0-7 A straight angle = 2 right angles = angle of a semicircle = sum of angles of a triangle.

All polygons can be built from triangles. To find the sum of all angles in a **polygon**, move along its perimeter until you return to the original point. This makes a 360° turn. If all the polygon's angles are the same and equal A, the turn at each vertex is $(180 - A)°$, as in Figure 0-8. With n sides and vertices the total turn is $n(180°\text{-}A) = 360°$, so A is given by Eqn. 0-3.

$$\boxed{A = 180^{\circ}\left(1 - \frac{2}{n}\right)} \qquad \text{Eqn. 0-3}$$

Eqn. 0-3 shows that for an equilateral triangle, angle $A = 180°(1\text{-}⅔) = 60°$, and for a square, angle $A = 180°(1\text{-}½) = 90°$ which you know. Eqn. 0-3 also shows that for a pentagon, angle, $A = 180°(1\text{-}{}^{2}/_{5}) = 108°$, which you may not have known or remembered.

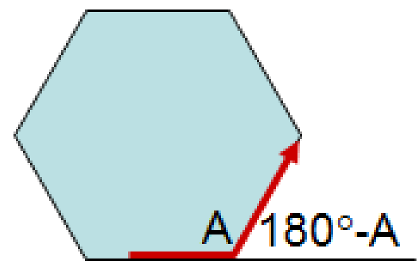


Figure 0-8 At each corner or vertex of a polygon, the direction along the perimeter changes by (180° - A).

Lengths and the Pythagorean Theorem: For Right Triangles, $c^2 = a^2 + b^2$

Without this ancient theorem (which probably predates Pythagoras by 1000 years) much of geometry, trigonometry, and Calculus would crumble. Pythagoras states, for a right triangle, the sum of the squares of the two side lengths touching the right angle ($a^2 + b^2$) equals the square of the hypotenuse length (c^2), the side opposite the right angle.

Proving the Pythagorean Theorem helps to keep it in your head and appreciate it. There are many proofs. I like the one that uses proportions and the geometry of similar triangles, illustrated in Figure 0-9. First, draw a right triangle with sides, a, b, and hypotenuse, c, and angles A, B, and C (the right angle). The key is to draw the (dashed) line from the right angle that is perpendicular to the hypotenuse, c. This creates two smaller right triangles that are both similar to the large right triangle, and divides the hypotenuse into two segments, x and $(c - x)$. What is x? Who cares? After it serves its purpose it magically disappears by the end of the derivation.

Because all three right triangles are similar, the ratios of the sides to the hypotenuse of each triangle are equal. Two resulting equations are:

$$\frac{a}{c} = \frac{x}{a} \quad \Rightarrow \quad a^2 = cx$$

$$\frac{b}{c} = \frac{c-x}{b} \quad \Rightarrow \quad b^2 = c^2 - cx$$

When these are added, x disappears and the result is the **Pythagorean Theorem**

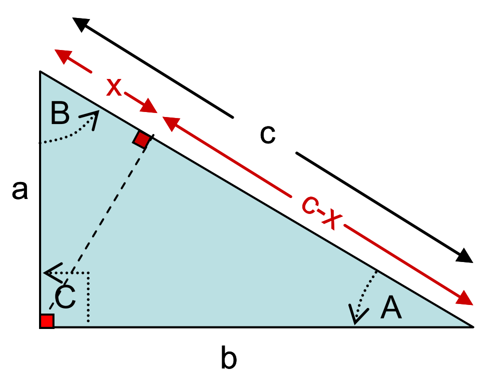
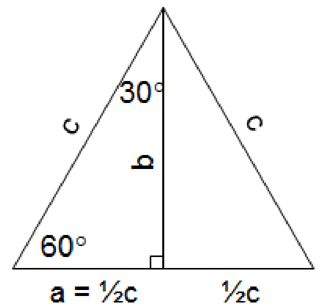


Figure 0-9 Graph for proving the Pythagorean Theorem

$$\boxed{a^2 + b^2 = c^2} \qquad \text{Eqn. 0-4}$$

Equilateral and Right Triangles

We see several triangles quite often. An Equilateral triangle has all 3 sides of equal length and all 3 angles equal to 60°. Cutting it in half (Figure 0-10) creates two right triangles with acute angles of 30° and 60°. The side opposite the 30° angle has half the length of a side of the original equilateral triangle or ½*c*. Then we find the length of the 3rd side (the height or altitude of the equilateral triangle) from the Pythagorean Theorem.



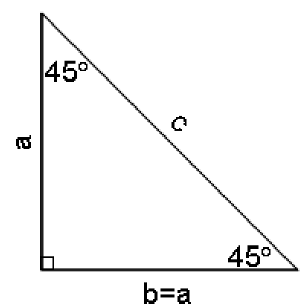
$$c^2 = a^2 + b^2$$

$$c^2 = \left(\frac{c}{2}\right)^2 + b^2$$

$$b = \sqrt{\frac{3c^2}{4}} = \frac{\sqrt{3}}{2}c$$

Figure 0-10 Dividing an equilateral triangle into two 30-60-90 right triangles.

An isosceles triangle has two equal sides and two equal angles opposite the equal sides. When the isosceles triangle is also a right triangle, as in Figure 0-11, each acute angle equals 45° and the lengths of the sides compared to the hypotenuse are solved below again using Pythagoras.



$$c^2 = a^2 + b^2$$

$$c^2 = 2a^2 = 2b^2$$

$$a = b = \frac{\sqrt{2}}{2}c$$

Figure 0-11 The isosceles right triangle.

0.3 Graphing and Cartesian Geometry: Axes to Grind and Grids to Iron

The Ancient Greeks did geometry without algebra. While Europe slept during the Dark Ages, Persian mathematicians, such as al-Khwārizmī (c. 780-850) and the poet, Omar Khayyam, (1048-1131) gave algebra the beginnings of its modern form and greatly sped calculations by replacing the old Roman numerals with Arabic numbers.

But geometry and algebra remained essentially separate branches of math until Descartes unified them in 1637 by overlaying geometrical figures such as rectangles and circles on a grid of squares, as in Figure 0-12. A pleasing legend is that Descartes got the idea while in bed watching a fly cross a ceiling of square tiles.

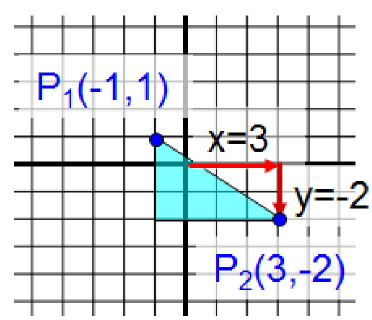


Figure 0-12 Cartesian grid with a right triangle.

On this grid (or map), horizontal and vertical distances are measured from a central point called the origin. The horizontal line through the origin is called the *x*-axis and *x* is the distance <u>to the right</u> of the origin. Any point to the left of the origin has a negative value of *x* (or negative longitude). Similarly, the vertical line through the origin is the *y* axis, and *y* is the distance above the origin. Any point below the origin has a negative value of *y* (or negative latitude). Putting this together, each point on the grid has a unique pair of *x* and *y* values, written (*x*, *y*). Thus a point 3 units to the right of the origin and two units below it (shown by the red arrows) has *x* = 3 and *y* = -2 or (*x*, *y*) = (3, -2).

This system makes a natural fit for applying the Pythagorean Theorem and calculating distances between two points. Figure 0-12 includes a (blue shaded) right triangle 3 units high and 4 wide. The Pythagorean Theorem shows that the hypotenuse or distance between Points P_1 and P_2 is 5 because $3^2 + 4^2 = 9 + 16 = 25 = 5^2$.

One of the great uses of the Pythagorean Theorem and Cartesian Coordinates occurs in Video Games. Animation is done by erasing objects and redrawing them in new nearby positions rapidly and repeatedly. Collisions, hits or kills in the Video Games occur when the distance between two objects is less than a certain tiny number so that the objects appear to collide. Then the programmer of the video game can make the objects bounce, merge, explode, disappear, die, etc. We are all seduced by these illusions because they can appear so realistic. I tell my students, "Don't waste your time playing video games! Make them!"

Axes to Grind

I have at least one ax(is) to grind. I cannot tell you how striking a contradiction it is for me to see college science students produce beautiful graphs using a program such as Excel but then not have any idea of the axes nor any ability to describe the graphs. I will attempt to fill this gap by literally putting flesh and blood in a graph.

Ht (in)	Wt (lbs)
19	7
25	15
62	97
69	170
73	280

Table 0-2 Heights and weights of 5 students in a class of 83.

Table 0-2 lists the heights and weights of 5 students in a class of 83 (to keep the table short). Each student's height and weight is plotted as a yellow or purple circle in Figure 0-13. The purple circles represent the 5 students listed in the table. The shortest (and lightest) student in the class was 19 inches tall and weighed 7 pounds. This was a precocious student – a newborn, who was the 2nd best student in the class (only kidding). A few other very young, precocious students were also in the class. The data point for the newborn is the lowest and furthest left circle in the Figure. The heaviest student was 280 pounds and 73 inches tall. Before reading further plot where you fit in Figure 0-13.

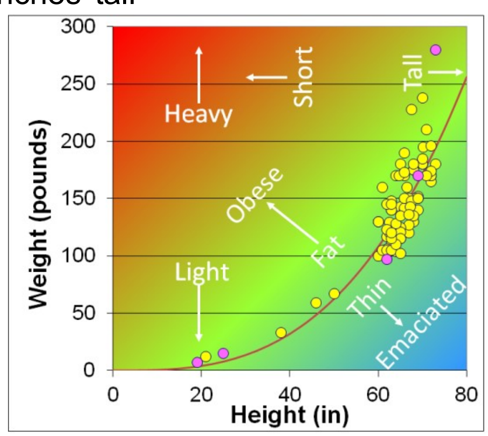


Figure 0-13 Heights and weights of 83 students in a college class. The purple curve represents an average human shape.

Let's discuss the main features of Figure 0-13. Height is plotted as the horizontal (*x*) axis. The further a data point is to the right the taller the person and the further a data point to the left the shorter the person.

Weight is plotted as the vertical (*y*) axis. The higher a data point on the graph the heavier or more massive the person while the lower a data point on the graph the lighter the person.

The purple curve is a cubic ($wt = k{\cdot}h^3$ where k is a constant). It shows the relationship between height, h and weight, wt for a person with an average shape. Why is the relation cubic? Weight or mass (weight = mass $\times$ gravity = $m{\cdot}g$) is proportional to volume, or length cubed. Any data point (person) above and to the left of the cubic curve is fatter than average (has a high body mass index). The region far above and to the left of the curve represents obesity.

Any data point below and to the right of the curve represents a person who is thinner than average. The region far below and to the right of the curve is the region of emaciation or anorexia – very light for a given height. Note that the newborn is very light but not skinny. Very tall people, for example NBA players, may well be thin even if they weigh over 200 lbs! Now think about your data point! If it lies far from the curve then this book has just taught you a major math lesson.

Slopes: Rates and Ratios (and Derivatives!)

Slope (often given by the letter, m) also called *gradient*, or *steepness*, is something you can see. In Calculus another word for slope is Derivative! We will meet it again and again!

The classic shape of a volcano such as Mount Fuji in Japan is a natural feature with a well-defined slope. Particles blasted out from the top come to rest at the angle of repose, the steepest angle that a particle will not slide or roll down. You can reproduce this by pouring salt, sugar, or dry sand onto a table, as in Figure 0-14. Each pile takes the shape of a circular cone with an angle of repose $\approx 35°$ and slope, $m \approx 0.70$. This is close to the slope of the hypotenuse of the 3: 4: 5 right triangle in Figure 0-12 ($m = 0.75$). You may be surprised that this is double the slope of the two steepest streets in San Francisco (m 0.315). Streets with a slope, $m = 0.315$ may sound like child's play until you try to drive or especially bike up or down them. Then they can be real scary.

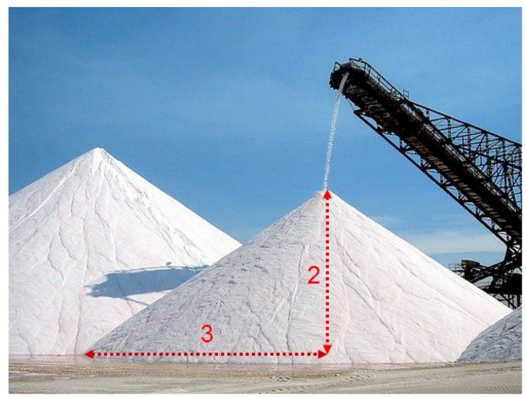


Figure 0-14. Pillars of salt grains form cones with a slope, $m \approx 0.70$, slightly more than 2/3.

Slope, m, is the **rise divided by the run**, or formally, the height difference between two points divided by the horizontal distance between the points (Eqn. 0-5).

$$\boxed{m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\Delta y}{\Delta x}} \qquad \text{Eqn. 0-5}$$

In Eqn. 0-5, the Greek symbol, Δ (Delta), is a prefix that means Δifference of or change of. You saw it in Table 0-1. Get used to it! It is universally used in Calculus.

Eqn. 0-5 gives slope a far more general meaning than simply the steepness of a hill. Slopes are ratios or rates. In Calculus, we call them Derivatives! This general meaning makes them exceedingly important in math, physics, economics and finance, chemistry, biology, forecasting, etc. Newton tackled the problem of how the planets move around the Sun. Their velocity is their rate of motion or the change of position divided by the change of time. Your salary is the hourly rate of income. Interest is typically an annual rate of increase of your savings or loans. In order to forecast any quantity such as temperature, we must know its rate of change. Derivatives are everywhere.

Look back to Figure 0-12 to calculate the slope of the hypotenuse of the triangle. At P_1, x_1 = -1 and y_1 = 1. At P_2, x_2 = 3 and y_2 = -2. Substituting into Eqn. 0-5 then yields

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{(-2)-1}{3-(-1)} = \frac{-3}{4} = -0.75$$

Note that the slope is negative! Negative and positive slopes may be equally steep, but there is a major difference. Imagine that the *x*-axis represents time, *t* and the *y*-axis represents your savings, *S*. In Figure 0-12 as time advances to the right your savings are going down. That is what a negative slope means, and it is a lot different than a line with a positive slope – i. e, one that rises toward the right, so that *S* increases with time.

In elementary algebra, you learn that the standard equation for a straight line is,

$$\boxed{y = mx + b} \qquad \text{Eqn. 0-6}$$

In Eqn. 0-6, *m* is the slope and *b* is the *y*-intercept or the value of *y* when *x* = 0.

Problem: Find the equation of the hypotenuse in Fig. 0-12.
Procedure: A: Since any two points define a line, choose two points on the line. The coordinates of points P_1 and P_2 are (-1, 1) and (3, -2). B: Substitute the *x* and *y* values of these points into Eqn. 0-6 and solve for *m* and *b*. Finally, C: confirm that *m* is the slope and *b* is the *y*-intercept.
Solution:

$$\begin{aligned} \text{at P}_1: \quad 1 &= m(-1) + b \\ \text{at P}_2: -2 &= m(3) \quad + b \end{aligned}$$

Subtracting (P_2-P_1) eliminates *b* and yields 4*m* = -3 or *m* = -¾. Substituting for *m* at P_1 yields 1+ *m* = *b* = ¼. Thus the equation of the hypotenuse in Figure 0-7 is,

$$y = -0.75x + 0.25$$

0.4 Trigonometry (Triangle Measure)

Much of Trigonometry amounts to fancy manipulations of the Pythagorean Theorem. Features of the right triangle in Figure 0-15 start with the longest side or hypotenuse (*hyp*), facing the right angle. The side touching both the acute angle, θ and the right angle is *adj* (adjacent), while the other side touching the right angle is *opp* (opposite).

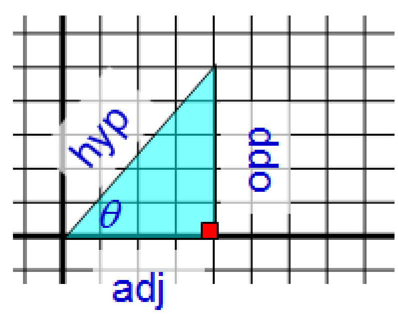


Figure 0-15 Sides and angles of the right triangle

Trigonometric Functions...

Trig functions are the 6 ratios of sides of right triangles in Eqn. 0-7. These ratios are important in math, physics, etc. For example, the ratio, *opp/adj is the* slope of the hypotenuse. It is called the tangent of the angle, θ, or tan(θ).

$$\begin{array}{lll} \cos(\theta) \equiv \dfrac{adj}{hyp} & \sin(\theta) \equiv \dfrac{opp}{hyp} & \tan(\theta) \equiv \dfrac{opp}{adj} \\ \sec(\theta) \equiv \dfrac{hyp}{adj} & \csc(\theta) \equiv \dfrac{hyp}{opp} & \cot(\theta) \equiv \dfrac{adj}{opp} \end{array}$$

Eqn. 0-7

Values of Trig Functions for Special Angles
0°, 30°, 45°, 60° and 90° are angles that are used as examples so often that it pays to memorize their sines, cosines, and tangents. The values, given in Table 0-3, are based on applying Pythagoras to Figures 0-10 and 0-11.

Angle (°)	Sin	Cos	Tan
0	0	1	0
30	0.500	0.866	0.866
45	0.707	0.707	1.000
60	0.866	0.500	1.155
90	1	0	±∞

Table 0-3 Values of sines, cosines and tangents.

Note: The Greek letter, θ is often used in Calculus for angles. Now is the time to begin to get used to Greek letters because you <u>will</u> see them very often!

Greek Letters – Get used to them!

In High School math (before Calculus), we were happy to use English letters such as *a*, *b*, *c* for constants, *x*, *y*, *z* for variables and *A*, *B*, *C* for angles. The only Greek letter I remember in High School math was π (pi). In Calculus, Greek letters spread like an epidemic that made learning Calculus even more difficult for all of us. But, get used to what they look like and what they sound like!

Why didn't they leave us alone with the good old English letters? The answer is simply because between math, physics, and chemistry, the English letters ran out. For example, time = *t*, Kelvin Temperature = *T*, so the only thing left for the period or the time to complete a cycle is the Greek letter, τ (tau).

In Calculus, Greek letters are often used for angles. The most popular angle is θ (theta), but there are also α, β, and γ, as well as ϕ, ψ, and sometimes λ for longitude, even though λ is more often used for wavelength.

A list of the small Greek letters is given below.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
α	β	χ	δ	ε	ϕ	γ	η	ι	φ	κ	λ	μ	ν	ο	π	θ	ρ	σ	τ	υ	ϖ	ω	ξ	ψ	ζ

In this book I use 10 lower case and 2 upper case Greek letters. Say them out loud as you read them: **α = alpha**, **δ = delta**, **ε = epsilon**, **ϕ = phi**, **λ = lambda**, **π = pi**, **θ = theta**, **ρ = rho**, **σ = sigma**, **τ = tau**, **Δ = Delta** and, **Σ = Sigma**.

Trigonometric Identities

Trigonometric identities are the relations between the trigonometric functions based on the Pythagorean Theorem. In Trig, you spent an eternity manipulating these and similar identities until you got sick and tired of them.

Perhaps the most fundamental of these identities uses Pythagoras to relate sines and cosines. Simply divide each term by hyp^2.

$$opp^2 + adj^2 = hyp^2 \quad \Rightarrow \left(\frac{opp}{hyp}\right)^2 + \left(\frac{adj}{hyp}\right)^2 = 1$$

Substituting the relevant trig functions of Eqn. 0-7 yields,

$$\boxed{\sin^2(\theta) + \cos^2(\theta) = 1} \qquad \text{Eqn. 0-8}$$

If, instead we divide Pythagoras by adj^2,

$$\left(\frac{opp}{adj}\right)^2 + \left(\frac{adj}{adj}\right)^2 = \left(\frac{hyp}{adj}\right)^2$$

And again substitute the relevant relations of Eqn. 0-7, the result is

$$\boxed{\tan^2(\theta) + 1 = \sec^2(\theta)} \qquad \text{Eqn. 0-9}$$

Some of the more sophisticated relationships that emerge from these exercises are the **law of cosines** (Eqn. 0-10), **law of sines** (Eqn.0-11), and multiple angle formulas, Eqns. 0-12, 0-13, and 0-14.

Deriving the Law of Cosines

Draw the left triangle in Figure 0-16. Divide it into two right triangles by creating altitude, y, so that we can use Pythagoras. This splits side, a into x and $a - x$. Eqn. 0-7 implies that $(a - x)/b = \cos(C)$. Rearranging,

$$x = a - b\cos(C)$$

Eqn. 0-7 also implies that,

$$y = b\sin(C)$$

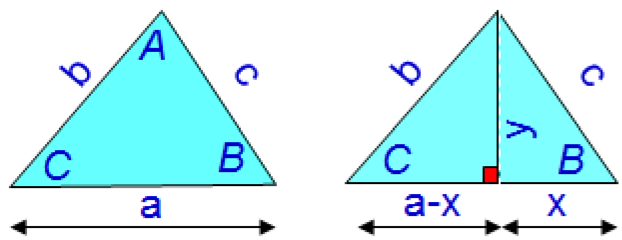


Figure 0-16 Draw an altitude to a triangle to prove the Laws of Cosines and Sines.

Then, use Pythagoras for the far right triangle, i. e., $c^2 = x^2 + y^2$, and replace x and y.

$$c^2 = x^2 + y^2 = a^2 + b^2\cos^2(C) - 2ab\cos(C) + b^2\sin^2(C)$$

$$c^2 = a^2 + b^2\left[\cos^2(C) + \sin^2(C)\right] - 2ab\cos(C)$$

Substituting Eqn. 0-8 then gives the Law of Cosines.

$$\boxed{c^2 = a^2 + b^2 - 2ab\cos(C)} \qquad \textbf{Eqn. 0-10}$$

The Law of Sines (and its Derivation)

Figure 0-16 also helps us derive the law of Sines. Looking at the two right triangles, you should see y equals both $c \cdot \sin(B)$ and $b \cdot \sin(C)$. Rearranging leads to,

$$c\sin(B) = b\sin(C) \quad \Rightarrow \quad \frac{\sin(B)}{b} = \frac{\sin(C)}{c}$$

To complete the trio, draw the altitude from vertex *C* to side *c*. You can then show that *a*·sin(*B*) = *b*·sin(*A*). This completes the relations of the Law of Sines,

$$\boxed{\frac{\sin(A)}{a} = \frac{\sin(B)}{b} = \frac{\sin(C)}{c}} \qquad \textbf{Eqn. 0-11}$$

You can use the Law of Sines to determine the lengths of 2 sides of any triangle without measuring if you know one length and the angles.

Multiple Angle Formulas

Trigonometric identities for multiple angles, (*A* + *B*) are given in Eqns. 0-12, 0-13a, and 0-14. Proving any of these requires the Pythagorean Theorem, but usually the key step of genius is to construct the right diagram.

$$\boxed{\sin(A+B) = \sin(A)\cos(B) + \cos(A)\sin(B)} \qquad \text{Eqn. 0-12}$$

$$\boxed{\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B)} \qquad \text{Eqn. 0-13a}$$

$$\boxed{\tan(A+B) = \frac{\tan(A)+\tan(B)}{1-\tan(A)\tan(B)}} \qquad \text{Eqn. 0-14}$$

Let's derive Eqn. 0-12. Try to draw Figure 0-17 yourself. First, draw the blue triangle for Angle, *A*. Second, draw the right triangle for angle *B* so that it sits on the blue triangle. Third, draw a right triangle for Angle (*A* + *B*) with the vertical line with length, *y* as the opposite side. Fourth, draw the yellow right triangle. Can you verify that its angle at top is equal to *A*? Here is the math!

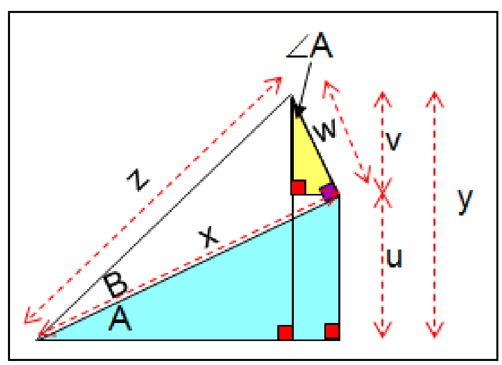


Figure 0-17 Graph constructed to prove the identity for sin(A+B).

1: Express sin(*A* + *B*) in terms of *u*, *v*, and *z*.

$$\sin(A+B) = \frac{y}{z} = \frac{u+v}{z}$$

2: Express *u* and *v* in terms of *z* and sines and cosines of Angles *A* and *B*.

$$\begin{array}{ll} u = x\sin(A) & v = w\cos(A) \\ x = z\cos(B) & w = z\sin(B) \\ u = z\sin(A)\cos(B) & v = z\cos(A)\sin(B) \end{array}$$

3: Substitute for *u* and *v* in the equation for sin(*A* + *B*).

$$\sin(A+B)=\frac{u+v}{z}=\frac{z\sin(A)\cos(B)+z\cos(A)\sin(B)}{z}$$

4: Simply, yes simply, cancel the *z*'s and the result is Eqn. 0-12.

Many books go on to give all the formulas for angle (*A* – *B*). This is a waste of space if you recall that cos(-*B*) = cos(*B*), sin(-*B*) = -sin(*B*), etc. However, one variant, the so-called half angle formula, is very useful for evaluating π, as we will see in Chapter 1.

Problem: Find sin(22.5°) and the half angle formula.
Information: 22.5° =½(45°) and we know sin(45°) = cos(45°) = $1/\sqrt{2}$.
Solution: Set angles *A* = *B* in Eqn. 0-13a, solve for sin(*A*).

$$\cos(2A)=\cos^2(A)-\sin^2(A)=1-2\sin^2(A)\Rightarrow\boxed{\sin(A)=\sqrt{\frac{1-\cos(2A)}{2}}}\quad \text{Eqn. 0-13b}$$

Eqn. 0-13b is the half angle formula for the sine! Now, substitute *A* = 22.5°.

$$\sin(22.5^\circ)=\sqrt{\frac{1-\cos(45^\circ)}{2}}=\sqrt{\frac{2-\sqrt{2}}{4}}\approx 0.383$$

Value of the Struggle

When I first tried to derive Eqn. 0-12, I bumbled around for an hour because I did not construct the right triangle for angle *B*. Then I gave up and looked it up (I used the Khan Academy). The diagram is always key, but the real key is the struggle. After my struggle, the moment I saw the correct diagram, I knew how to derive the law. The moral:

To learn any subject, struggle first. Look up the solution only after you have tried to get it and can't. Your struggle teaches you more than you can imagine. But, get to know your optimal struggling time. Too little and you won't learn. Too much and you waste time.

Burning Calories

By now, you must be exhausted. I know I am. Thinking burns far more calories than you might think. Our brains weigh only about 1.4 kg – about 2% of total body weight but they burn about 20% of the calories we consume when we are at rest, about 260 per day. Hard thinking burns calories at a much higher rate. In one experiment, two groups of students were compared. One had taken a difficult test for about 45 minutes. The other group had been relaxing. Then all were led to the cafeteria. The students who took the test ate an average of 200 calories more. Hence the expression, "my brain is fried."

Why don't you burn some calories thinking about measuring your waist! You can do it with a tape measure. But if you only have a ruler, you can measure the diameter of your waist and then multiply by about 3.14. This assumes that your waist is shaped like a circle. Actually, your waist is shaped more like an ellipse, so I would not use this way of estimating your waist to buy clothes. But it's not a bad start.

0.5 Geometry of the Circle

The Ancient Greeks thought that the circle is the perfect shape. Circles lie at the foundation of technology, from the wheel to the gear. Circles involve cycles, and we will see that they are related to waves. The Greeks also thought that the orbits of the planets and the Sun are circles, but they were wrong – the orbits are ellipses – slightly squashed circles. Nevertheless, for more than 2000 years it was heresy to suggest that heavenly orbits were not circles.

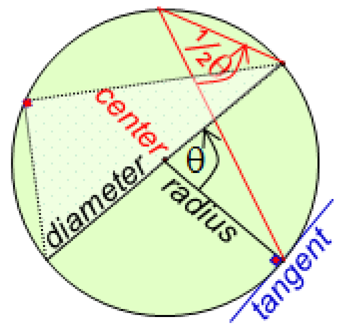


Figure 0-18 Properties of circles.

A circle is the curve that includes all points on a flat plane located at a fixed distance, called the radius, *r*, from a central point. You can draw a circle with a compass by fixing the sharp point on the paper and spinning the other arm with a pencil a set distance away. A sphere is the solid that includes all points out to a fixed distance, called the radius, *r*, from a central point in all directions. The sphere is the 3-dimensional version of the circle. The edge of any straight cut through a sphere is a circle.

Here are some of the marvelous properties of circles (see Figure 0-18) and spheres you should never forget.

1: A tangent line to the circle is perpendicular to the radius at the point of contact.

2: An inscribed angle (with its vertex on the perimeter) is half the central angle (with a vertex at the center) that subtends the same arc. Thus, a triangle inscribed in a semicircle is a right triangle.

3. The circle's width or height, diameter, *d* is twice the radius

$$\boxed{d = 2r} \qquad \text{Eqn. 0-15}$$

4: The length of the perimeter of the circle, i. e., the circumference, $C = 2\pi r$. Measure both C and d with a tape measure and then calculate the ratio just to satisfy yourself. You may not get the exact answer, but you will be close.

$$\boxed{C = 2\pi r = \pi d} \qquad \text{Eqn. 0-16}$$

6: The magical number, π appears again in Eqn. 0-17 for the area of a circle, $A_{\text{circle}} = \pi r^2$.

$$\boxed{A_{\text{circle}} = \pi r^2} \qquad \text{Eqn. 0-17}$$

7: The magical π continues to appear when the shape is a sphere. (Note the poem!)

$$\boxed{A_{\text{sphere}} = 4\pi r^2} \qquad \text{Eqn. 0-18}$$

This extraordinary relation, namely that surface area is exactly 4 times greater for a sphere than for a circle with the same radius, means that average sunlight intensity on any planet is ¼$^{\text{th}}$ of direct sunlight intensity. This relationship determines average Temperature on every planet including Earth, and is built into every climate model.

8: The volume of a sphere also involves π (and *r*), and is,

$$\boxed{V_{\text{sphere}} = \frac{4}{3}\pi r^3} \qquad \text{Eqn. 0-19}$$

Because π is irrational, all attempts to square the circle (draw a square with the same area as a circle), which even occupied Abraham Lincoln, are doomed to fail. I remember the argument in grade school. Is π = 3.14 or 31/7? We lined up in battle formation to defend our ignorance and errors. Another war for nothing – but at least no one died!

In Chapter 1, we will see how the Ancients closed in on the value of π and the area of a circle. Then in Chapter 6, we will use Calculus to prove Eqn. 0-18 and Eqn. 0-19 for a sphere's surface area and volume. It will take a long time to get there but the proofs are amazingly short.

A Natural Measure for Angles: Radians for Degrees or New Lamps for Old

In the story of Aladdin, the evil Magician gave new lamps for old to get the magic lamp. In our story – Calculus – the old angles are degrees and the new angles are radians. But here, the radians are magic. The degree is only good because we are so used to it. What if we were all living on Mars where the year is 687 days? Martians would probably give the circle 690°. The radian is superior because it defines angle as the ratio of the arc's length, *s*, to the radius, *r* (as in Figure 0-19).

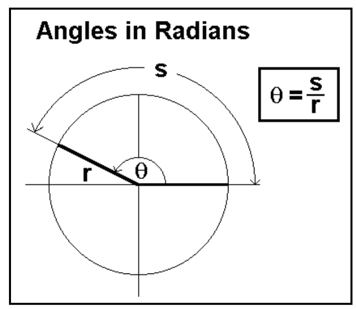


Figure 0-19 θ and *s* start at 3:00 and increase counterclockwise.

When the unit for angle is radians, distance along the circumference is,

$$\boxed{s = r\theta} \qquad \text{Eqn. 0-20}$$

One radian is the angle of a circle's wedge for which *s* = *r*. A circle has 2π radians, no matter what planet you're on. This makes radians ideal for calculating length and area.

The Circle, Cartesian Coordinates, and Trigonometry

Great things happen when the circle is superimposed on a Cartesian Graph, as in Figure 0-20. We obey standard convention by starting on the *x* axis, at 3:00 o'clock, and move counterclockwise. As *r* rotates counterclockwise from the x-axis by an angle, θ, it sweeps out a distance, $s = r\cdot\theta$ along the circumference. At any instant, it forms a triangle whose height, $y = r\cdot\sin(\theta)$ and whose distance to the right of the origin is $x = r\cdot\cos(\theta)$. (This leads to the circle's equation, Eqn. 0-24!)

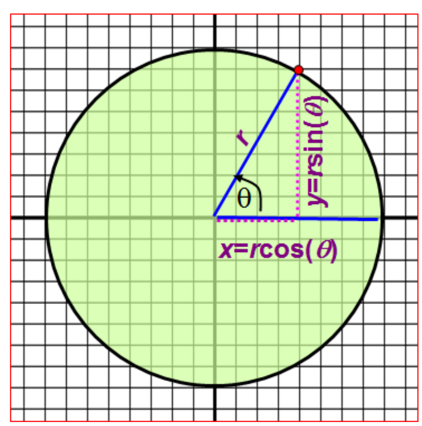


Figure 0-20 The circle, angle, θ, and its sines and cosines.

This suggests that the sine is a lot more than the ratio, *opp/hyp* for a right triangle. If you persist in thinking of the sine merely as the ratio, *opp/hyp* for a right triangle, then θ can never be more than $\pi/2$. However, if you think in terms of the clock, or in terms of winding a rope around a dear friend's waist, then angle, θ can increase forever. At one complete cycle, $\theta = 2\pi$ radians (360°). After 2 complete cycles, $\theta = 4\pi$ radians, etc.

The sine is then like the height of the hand of a clock in Figure 0-21 with $r = 1$ and whose center is at the origin. The hand starts at 3:00, or $(x, y) = (1, 0)$. Then it turns *counter*clockwise. At first the hand rises rapidly but by the time it reaches 12:00, it levels off at the top. At that point, angle θ is $\pi/2$ radians (or 90°). As the hand continues, it begins to descend but θ continues to increase.

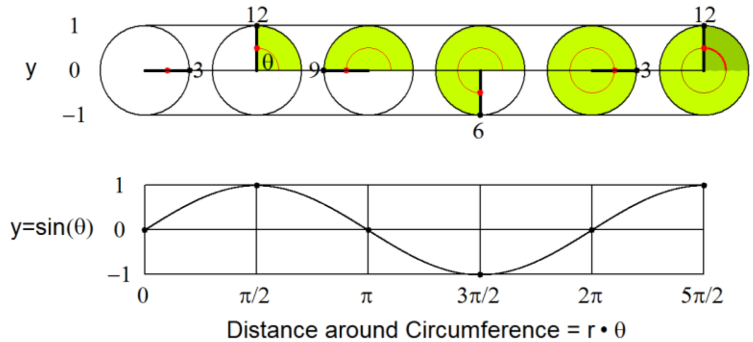


Figure 0-21 Sine (and cosine) waves result when the red dot moves around the circle.

The bottom graph of Figure 0-21 shows the sine curve or the height of the clock hand as a function of the distance around the circle or the angle, θ. What is so wonderful about the sine curve? It is a wave, and waves are very common and important in nature. There are sound waves, electromagnetic waves, water waves, and (in a few cases) even brain waves. Some of our most intimate motions are oscillatory.

Inverse Trigonometric Functions

In the beginning we learned about angles. Then, despite their difficulty, we had to learn the trigonometric (trig) functions, tangent, sine, and cosine. Why? Because nature so often gives us the trig functions! Often we know the slope and not its angle. The tangent is the slope. We find waves everywhere in nature and the solutions to equations describing waves are usually given in terms of sines and cosines rather than the angles.

So, we often find ourselves in the uncomfortable position of knowing the trig function and trying to determine its angle. That is difficult and may seem obscure, even mystifying or terrifying, but it is actually like going back home to the angles we knew in the first place. But how do you get back home to the angle?

The inverse trigonometric functions are simply angles. If, for example, you know the slope of a hill (i.e., its tangent) then you can measure its angle (or inverse tangent) with a protractor or find it by hitting $\tan^{-1}$ on a calculator. Since the trig functions and inverse trig functions are reciprocal, their relations are illustrated for the tangent in word form as

You Know the Angle	→	Use tan	→	Find the Tangent
You Know the Tangent	→	Use $\tan^{-1}$	→	Find the Angle

In equation form this appears as,

$$y = \tan(\theta) \quad \Leftrightarrow \quad \theta = \tan^{-1}(y)$$

Problem: Find the angle, θ, of the slope of the steepest street in San Francisco.
Information: The slope is 0.315 so that $\tan(\theta) = 0.315$.
Solution: Take the inverse tangent

$$0.315 = \tan(\theta) \quad \Leftrightarrow \quad \theta = \tan^{-1}(0.315) = 17.5^{\circ}$$

By now you have noted that I wrote the inverse tangent of *y* as $\tan^{-1}(y)$. The inverse trig functions of ***y*** are also written in this book as **$\cos^{-1}$(y), $\sin^{-1}$(y),** and**, $\tan^{-1}$(y)** as well as on most scientific calculators. This notation is common but troublesome because...

Warning: $\sin^{-1}$(y) DOES NOT MEAN 1/sin(y)! In equation form,

$$\sin^{-1}(y) \neq \frac{1}{\sin(y)}$$

Note: In an attempt to avoid this potential and likely confusion, the inverse trig functions are sometimes written as arccos, arcsin and arctan (or acos, asin and atan in computer programming languages). The arc is the arc (or angle) of a circle.

Problem: Find $\sin^{-1}(0.5)$
Solution: Figure 0-10 shows that the angle whose sine is 0.5 is 30°. It also appears in Table 0-3 and, of course, you can use a calculator.

Trick Problem: Find $\sin^{-1}$(1.25)
Solution: The maximum value of the sine is 1.0. Therefore, no angle has a sine = 1.25.

Warning: Inverse Trig Functions Have Many Values and hence must be defined over Principal Domains. Because the trig functions are waves, they repeat every 360°. Thus, for example, there are an infinite number of angles whose sine is 0.5, starting with 30° and 150°. Furthermore, the tangent is discontinuous every 180°. As a result the inverse trig functions are only guaranteed to be continuous and not repeat (i. e., be single-valued) over limited domains. The principal domains (and those given by the calculator) are defined to be -90° to +90° for $\sin^{-1}$ and $\tan^{-1}$ and 0° to 180° for $\cos^{-1}$.

0.6 Algebra and the Quadratic Formula

Quadratic equations are just about the only nonlinear equations that can be solved analytically, and that is fortunate because they are both very important and very common. The Pythagorean Theorem is quadratic. So too are the equations for curves that include circles and parabolas (Section 1.7). The solution to the quadratic equation,

$$\boxed{ax^2 + bx + c = 0} \qquad \text{Eqn. 0-21}$$

is the famous **Quadratic Formula**,

$$\boxed{x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}} \qquad \text{Eqn. 0-22}$$

The quantity in the square root ($b^2 - 4ac$) is the discriminant.

It is valuable to present for your viewing pleasure the steps in **The Derivation of the Quadratic Formula**.

1: Divide Eqn. 0-21 by *a*.

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

2: Complete the square of terms involving *x* by subtracting c/a and adding [*b*/2*a*]^2 to each side. Then, the left and right hand sides become,

$$x^2 + \frac{b}{a}x + \left(\frac{b}{2a}\right)^2 = \left(x + \frac{b}{2a}\right)^2 \qquad \left(\frac{b}{2a}\right)^2 - \frac{c}{a} = \frac{b^2 - 4ac}{(2a)^2}$$

Thus the useful form of the equation becomes

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{(2a)^2}$$

3: Take the square root of each side and subtract *b*/(2*a*) from each side. This yields the Quadratic Formula, Eqn. 0-22.

Testing the Quadratic Formula

There is an easy way to test difficult equations and formulas. Start with a solution that you have made up! Then see if the formula reproduces the solution. For example, say we know two solutions to *x*, namely, *x* = ¾ or 4*x* = 3 or, (4*x* – 3) = 0 and *x* = -4 or, (*x* + 4) = 0. We can multiply them into one equation to get,

$$(4x - 3)(x + 4) = 0$$

If we multiply this out, it becomes the quadratic equation,

$$4x^2 + 13x - 12 = 0$$

The coefficients, *a*, *b*, and *c* are then

$$a = 4 \qquad b = 13 \qquad c = -12$$

Substituting into the quadratic equation yields the desired results.

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-13 \pm \sqrt{13^2 - 4(4)(-12)}}{2(4)} = -\frac{13 \pm 19}{8}$$

Thus, we confirm the solutions we already knew,

$$x_1 = \frac{6}{8} = \frac{3}{4} \qquad x_2 = -\frac{32}{8} = -4$$

0.7 Quadratic Equations and Curves: Conic Sections and Analytic Geometry

Three demonstrations impressed me greatly as a student in Plane Geometry. First, the teacher, Mr. Richards placed a yardstick vertically against the corner of the blackboard and drew a line. Then he tilted the yardstick in steps by small angles, each time drawing another line. Figure 0-22 shows that the envelope of straight lines gives a powerful illusion of a curve. At the time I had no idea that it gave a sneak peek at Calculus, because in a way, Calculus is the math of curves.

For the second demonstration, Mr. Richards began by pressing a thumb tack into a piece of wood. Then he placed a loop of string around the tack and stretched the loop with a pencil. Keeping the loop taut, he moved the pencil around the tack, and it traced out a perfect circle. That was no surprise. But then he pressed a second tack into the wood, and stretched the loop around both tacks. As he moved the pencil around, keeping the loop taut, it traced out a perfect ellipse. That was a surprise! Years later I learned that you can draw a parabola and a hyperbola as well as a circle and an ellipse using only string, tacks, a ruler, and a pencil.

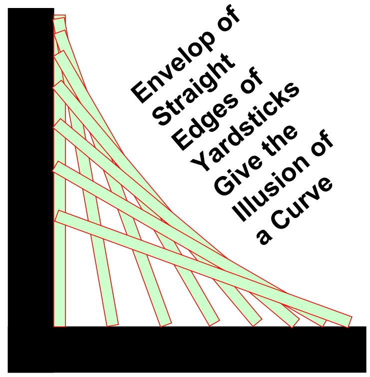


Figure 0-22 An envelope of tilted straight edges appears to be a curve.

The third demonstration (Figure 0-23) used a wooden cone to show that a slice straight through a cone, called a conic section is either a circle, ellipse, parabola, or hyperbola depending on the angle of the slice. When the slice is horizontal, it forms a circle (red). When the slice slopes less than the left edge of the cone, it forms an ellipse (green). A slice parallel to the edge of the cone is a parabola (blue). Finally, a slice with a greater slope than the edge of the cone (brown) is a hyperbola. These curves are graphed in Figure 0-24.

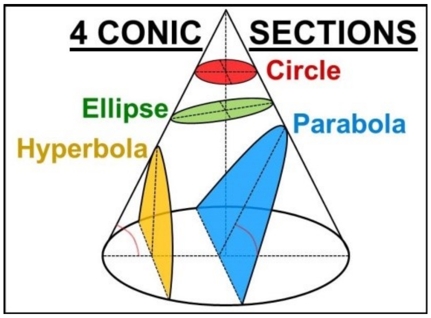


Figure 0-23 Conic sections.

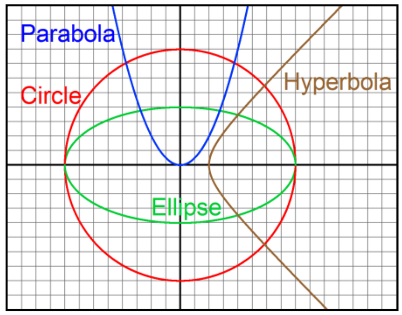


Figure 0-24 The conic curves.

The conic sections form the simplest family of curves because they are represented by quadratic equations - the simplest **non**linear equations. If the curves are rotated so that they are symmetrical about the *x* or *y* axes, the general quadratic equation is,

$$\boxed{Ax^2 + By^2 + Cx + Dy + E = 0} \qquad \text{Eqn. 0-23}$$

The relations between the constant coefficients, *A* and *B* in Eqn. 0-23 and listed in Table 0-4 decide which curve you get. For example, a circle results when *A* = *B*, and a parabola results when either *A* = 0 or *B* = 0.

Relation of Coefficients	Curve
$A = B$	Circle
$AB > 0$ ($A \neq B$)	Ellipse
$AB = 0$	Parabola
$AB < 0$	Hyperbola

Table 0-4 The conic section coefficients.

The conic sections occur everywhere in nature and technology. The circle has been called the perfect curve. It is the basis of the wheel and the gear, just as the sphere is the basis of the ball. Without the circle and sphere the world would wobble, bump, and grind to a halt.

The Circle and Ellipse

All points of a circle are the same distance, *r*, from a central point or focus, where *r* is the radius. The circle's equation is,

$$\boxed{\frac{x^2}{r^2} + \frac{y^2}{r^2} = 1} \qquad \text{Eqn. 0-24}$$

The standard equation for an ellipse is,

$$\boxed{\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1} \qquad \text{Eqn. 0-25}$$

An ellipse is a squashed or stretched circle. The height of an ellipse with the same width as a circle ($a = r$) is b/a times the circle's height at every value of x.

The ellipse is a cosmic curve - the orbits of the planets are ellipses. A circle has one focus but an ellipse has two. A marvelous property of the ellipse, shown in Figure 0-25, is that the total distance from one focus to any point (x, y) on an ellipse to the other focus is constant. This is why it is drawn with a stretched loop of string around a pencil and the foci.

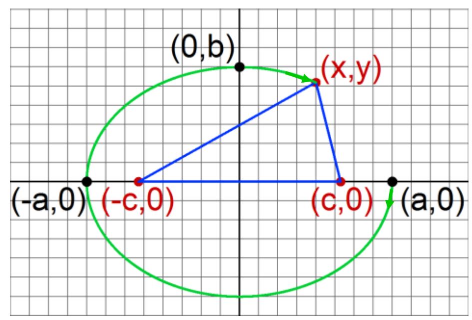


Figure 0-25 Drawing an ellipse with a loop of blue cord around the foci.

The ellipse of Eqn. 0-25 is centered at the origin, the y and x axes are axes of symmetry, the highest point is $y = b$, the rightmost point is $x = a$. When the ellipse is squashed vertically (in the y direction), $b < a$, the two foci (or thumb tacks) are located at points, ($\pm c$, 0), the total distance from the foci to any point on the ellipse = $2a$, and,

$$\boxed{c^2 = a^2 - b^2} \qquad \text{Eqn. 0-26}$$

Eccentricity, ε is the standard measure of how squashed an ellipse is. It is given by

$$\boxed{\varepsilon = \sqrt{1 - \frac{b^2}{a^2}} = \frac{c}{a}} \qquad \text{Eqn. 0-27}$$

There are two degenerate cases of the ellipse. When the ellipse rounds to a circle $b = a$, so $c = 0$, and $\varepsilon = 0$; when it flattens to a line, $b = 0$, so $c = a$, and $\varepsilon = 1$ (its maximum).

Problem: Find a, b, c, and ε of Earth's orbit around the Sun.
Information: Earth's obit is an ellipse, with the Sun at one of the foci. (The other focus still exists but nothing is there). Its closest (perihelion) and furthest (aphelion) distances to the Sun are $147(10)^8$ km and $152(10)^8$ km and they occur half a year apart (on the x axis in Figure 0-25).
Solution: Assume the Sun is fixed at focus, ($-c$, 0). Then, distances at aphelion and perihelion on the opposite side of the ellipse are,

$$\begin{array}{l} a + c \approx 152(10)^8 \\ a - c \approx 147(10)^8 \end{array} \quad \Rightarrow \quad c \approx 2.5(10)^8 \quad \Rightarrow \quad a \approx 149.5(10)^8$$

Eqn. 0-26 gives b and Eqn. 0-27 gives eccentricity, ε.

$$b = \sqrt{a^2 - c^2} \approx \sqrt{149.5^2 - 2.5^2} \approx 149.48 \qquad\qquad \varepsilon = \frac{c}{a} \approx \frac{2.5}{149.5} \approx 0.0167$$

The Parabola

The trajectory of a projectile not slowed by friction is a parabola (see Figure 5-1). A parabola rotated about its axis forms a paraboloid. That is the shape of the surface of water in a spinning beaker and of parabolic mirrors and antennas. Telescopes and solar heaters use parabolic mirrors because parabolas focus distant light and sound onto a central point appropriately called the focus when the parabola is pointed directly toward the source of the distant light or sound. For the same reason the ears of foxes and rabbits curve like parabolas.

Just as the circle consists of all points an equal distance from a central focus point, the parabola consists of all points an equal distance from the focus and a line (called the <u>directrix</u>, which is perpendicular to the parabola's axis of symmetry). It is customary to place the focus at (x, y) = (0, p) and the directrix at $y = -p$, as in Figure 0-26. This makes the y axis the axis of symmetry. Since lines A and A' have the same length (distance from the focus to the parabola = distance from the parabola to the directrix), Pythagoras tells us that

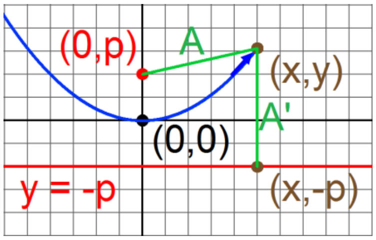


Figure 0-26 Parabola, $x^2 = 4py$ with focus at (0, p) axis of symmetry (x = 0) and directrix (y = 0, red line). Green lines, A, and A' are equal in length.

$$(y-p)^2 + x^2 = (y+p)^2$$

Expanding and cancelling equal terms leads to the standard equation for the parabola,

$$\boxed{x^2 = 4py} \qquad \text{Eqn. 0-28}$$

Problem: Find the location of the focus for the parabola $y = x^2$.
Solution: Solve Eqn. 0-28 for p. The focus is located at point (0, p).

$$x^2 = y = 4py \quad \Rightarrow \quad p = 0.25$$

We are now ready to prove that a parabolic mirror will focus distant light or sound at its focus. To do this we must first find an equation for the parabola's slope. Then, we must show if, in fact, distant light or sound is focused at the focus it obeys the **Law of Reflection**, namely, that the angle of Incidence, i equals the angle of reflection, r. This is shown in Figure 0-27.

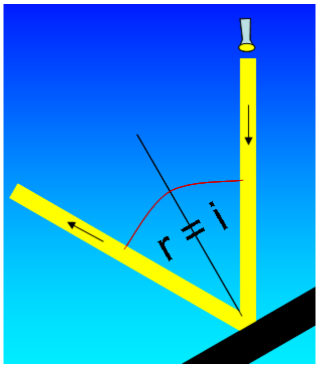


Figure 0-27 Law of Reflection: i = r.

Problem: Find the slope of the parabola given by Eqn. 0-28.
Solution: In Chapter 2, we will find the slope of many curves using Calculus. The extraordinary thing is that we can find the slope of a parabola using algebra because the quadratic formula provides an analytic solution. The equation of a line that touches $x^2 = 4py$ is $y = mx + b$. Combining these equations yields what seems to be a quadratic equation for x.

$$y = mx + b = \frac{x^2}{4p} \quad \Rightarrow \quad \boxed{\frac{x^2}{4p} - mx - b = 0} \qquad \text{Eqn. 0-29}$$

But surprise! Instead of solving Eqn. 0-29 for *x*, we use it to solve for slope, *m*. The ingenious step is that a line tangent to a parabola touches it at only one point. And, a quadratic equation has one solution only when its discriminant equals zero. Setting the discriminant of Eqn. 0-29 equal to 0, yields the equation relating *m* and *b*,

$$b = -pm^2 \qquad \text{Eqn. 0-30}$$

Next, replace *b* with $-pm^2$ in Eqn. 0-29. Rearranging shows that it can be treated as a quadratic equation for *m* in terms of x!

$$pm^2 - xm + \frac{x^2}{4p} = 0$$

Solving using the quadratic formula yields the unique solution for the slope, *m* of the parabola, $x^2 = 4py$ in terms of *x*!

$$\boxed{m_{x^2=4py} = \frac{x}{2p}} \qquad \text{Eqn. 0-31}$$

Squaring Eqn. 0-31 and replacing *m* with *b* from Eqn. 0-30 yields the equation of *b* for the tangent line to the parabola in terms of *x*.

$$\boxed{b = -\frac{x^2}{4p}} \qquad \text{Eqn. 0-32}$$

Now, we are ready to solve the problem…

Problem: Prove that a parabolic mirror aimed toward a distant light beam reflects it to the parabola's focus.

Solution: The proof depends on constructing the proper diagram. In Figure 0-28, draw the parabola, $x^2 = 4py$ with focus at point, (0, *p*). Next, draw the yellow light beam parallel to the *y* axis that strikes the parabola at point, (*x′*, *y′*) and reflects to the focus. The two wide segments of the light beam have the same length when the vertical segment extends up to point, (*x′*, 2*y′*+*p*). Draw the purple dashed line segment, Line, #2 connecting the focus (0, *p*) to the point, (*x′*, 2*y′*+*p*) on the light beam. Note: Because we made the two wide yellow segments equal, they and Line #2 form an isosceles triangle. Thus, the two angles facing the equal sides are equal. Finally, draw the purple dashed line tangent to the parabola where the light beam first struck it.

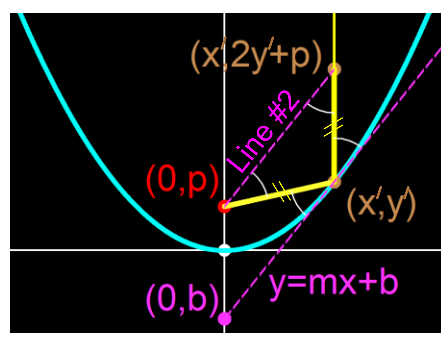


Figure 0-28 Diagram used to prove that a parabola reflects distant light so that it passes through the focus.

This elaborate construction solves the problem provided we can prove that the two purple dashed lines are parallel, for then the 4 angles marked with the white circle arcs are all equal and thus the light beams that pass through the focus obey the Law of Reflection.

To prove that Line #2 is parallel to the tangent line, we must show that its slope is equal to *m* from Eqn. 0-31. Using the equation for slope, Eqn. 0-5, Figure 0-28, and the equation of the parabola, Eqn. 0-27 yields,

$$m_{Line\#2} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{(2y'+p) - p}{x'-0} = \frac{2y'}{x'} = \frac{2x'^2}{4px'} = \frac{x'}{2p}$$

Thus, line #2 has the same slope as the tangent line so both lines are parallel and the proof that parabolas focus distant sound and light at the focus is complete. Parabolic reflectors are so effective that light reflected to the focus can start a fire, and you can hear people whispering a block away if you keep your ear near the focus.

The Hyperbola

Hyperbole is exaggerated talk, so it is no surprise that a hyperbola is an exaggerated curve. It is the path of a rocket that has escaped Earth's gravity after it has exhausted its fuel. It is also the curve of the shadow of the Sun on level ground in places where the Sun sets and hence, the curve etched into sundials outside the Polar Regions. It is also useful in signal detection as we will soon see. Its standard equation is,

$$\boxed{\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1} \qquad \text{Eqn. 0-33}$$

This equation describes two ‘dueling’ hyperbolas, for example, the two yellow curves in Figure 0-29 – one that opens in the plus *x* direction and one that opens in the minus *x* direction. Eqn. 0-33 looks like the equation for the ellipse (Eqn. 0-25) except that in front of the y^2 term it has a minus sign, whereas the ellipse has a plus sign. That difference of sign makes all the difference in the world, for the ellipse is the curve of a bounded orbit, while the hyperbola is unbounded. The two hyperbolas in Figure 0-29 are separated by distance, 2*a*, and their foci are located at $x = \pm c$, where *c* is given by,

$$\boxed{c^2 = a^2 + b^2} \qquad \text{Eqn. 0-34}$$

Note: For hyperbolas, $c > a$: for ellipses, $c < a$.

Also, while the ellipse is the curve that results when the sum of the distances to each focus is constant = 2***a***, the hyperbola is the curve that results when the difference of the distances to each focus is constant = 2***a***.

This property makes the hyperbola valuable in signal location. Simple detectors record the time (but not the direction from which) a signal arrives. The time lag for a signal travelling at the speed of light translates to a difference of distance, between two detectors. This defines a hyperbola.

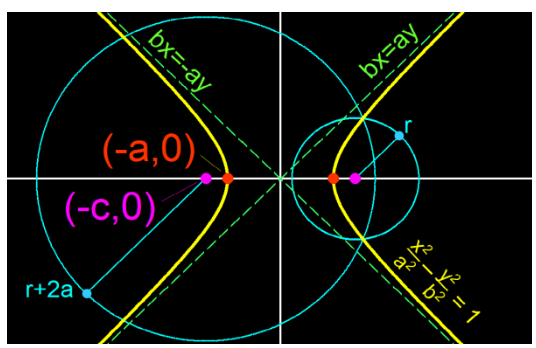


Figure 0-29 Hyperbolas (yellow curves) are all the intersecting points of pairs of the blue circles centered at opposite foci ($x = \pm c$). Green dashed lines are asymptotes.

With a 3rd detector we can generate two more hyperbolas. The intersection point of the hyperbolas pinpoints the source.

Problem: A secret agent sends out a signal for help. It arrives 1/30,000th s earlier at Detector #1, at (x, y) = (13 km, 0) than at Detector #2, at (-13 km, 0). Given the speed of light, $3(10)^5$ km·s^{-1}, the signal originated 10 km closer to Detector #1. Find the equation of the hyperbola on which the signal was sent.
Solution: Each detector is located at a focus, so ***c* = 13**. Since Detector #1 is 10 km closer to the signal then Detector #2, 2a = 10, so ***a* = 5**. Eqn. 0-34 then gives b.

$$b^2 = c^2 - a^2 = 13^2 - 5^2 = 12^2$$

Therefore, the equation of the hyperbola is,

$$\frac{x^2}{5^2} - \frac{y^2}{12^2} = 1$$

Figure 0-29 depicts two interesting features of hyperbolas. The first is linked to signal detection. Draw a blue circle with radius, r, centered at focus, (c, 0) and another with radius, $r + 2a$, centered at the other focus, ($-c$, 0). Then for every $r > c - a$, there are two points where the circles intersect. As r increases, these points trace out a hyperbola.

Second, as the magnitude of x in Eqn. 0-33 gets very large, so does the magnitude of y. Therefore, at great distances from the origin, we can neglect 1 because both $(x/a)^2 >> 1$ and $(y/b)^2 >> 1$, and the hyperbola can be approximated by the two straight lines,

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} \approx 0 \quad \Rightarrow \quad \boxed{y \approx \pm\frac{b}{a}x} \qquad \text{Eqn. 0-35}$$

These straight lines are called the asymptotes because the hyperbola approaches them asymptotically as |x| and |y| get large, but never quite reaches them.

The hyperbola also has its own eccentricity, given by,

$$\boxed{\varepsilon = \sqrt{1 + \frac{b^2}{a^2}} = \frac{c}{a}} \qquad \text{Eqn. 0-36}$$

Note that the hyperbola's eccentricity is always greater than 1, except when b = 0, but then the curve is a parabola!

0.8 Functions, Variables, and Dependence

In algebra we were relatively simple minded and relatively happy. We said things like, $y = x^2$. This is the…

Old Way $$y = x^2$$

In Calculus, we often get fancier. We take the same equation and say that x^2 is a ***function*** of x, or, x^2 equals f of x, or f(x). In equation form, we write in the…

New Way $$f(x) = x^2$$

Function is a word we use all the time. Your salary is a function of the hourly rate and the number of hours you work. In this sense, function of means depends on. Function has the same practical meaning in most branches of science. The speed of a rock you have dropped depends on (i. e., is a function of) how far or how much time it has fallen. Atmospheric pressure is a function of the height above sea level. Blood pressure is a more complicated function of weight, age, genetics, stress, and physical activity.

Functions can represent processes or the output of complex systems such as cells, plants, animals or governments. Plant growth and crop yield are functions of sunshine, temperature, rainfall, soil nutrients, etc. Viewed from inside, systems are complex entities of interacting components. Viewed from outside, systems are like black boxes that have an input, x, and an output, f(x) that is some function of x.

We are free to enter any input. For example, within limits, we can eat as many hot dogs as we want. Since we are free to choose the input, we call it the independent variable. But the output is a function of the system that depends on the input, so we call it the dependent variable. In this sense, many systems can be modeled and solved mathematically, as we will do in Chapters 1 and 5.

Function Games: Functions of Functions

Give mathematicians an inch and they'll take a mile. The moment they have defined functions, they will 1: brag about their generality by using strange letters and, 2: start talking about functions of functions. Here are two example of what they might do.

Problem: If f(x) = x^2, find f(u)!
Solution:

$$f(x) = x^2 \quad \Rightarrow \quad f(u) = u^2$$

Wow, that was so simple you can't believe it is the answer.

Problem: How do we express a function, f(u) = u^3 in terms of x if u = g(x) = sin(x)?
Solution: We must find f of g of x or f(g(x)).

$$f(u) = u^3 \quad \& \quad g(x) = \sin(x) \quad \Rightarrow \quad f\big(g(x)\big) = \big(\sin(x)\big)^3$$

All this is fine, but what is its use? When do we need functions of functions? Answer: any time cause and effect is complicated, i. e., whenever A depends on B but B depends on C. For example, your salary depends on the hourly rate and the number of hours you work. But both the hourly rate and the number of hours depends on your job and your boss, and the number of hours you work also depends on how interesting the job is, how desperate or driven you are and how early you leave work to learn Calculus.

Alert: Reversing the order of functions often produces a different result. [-(x^2) ≠ (-x)^2.]

$$g(u) = \sin(u) \quad \& \quad f(x) = x^3 \quad \Rightarrow \quad g\big(f(x)\big) = \sin\left(x^3\right) \neq \sin^3(x)$$

$$x = 1 \quad \Rightarrow \quad \sin(1^3) = \sin(1) \approx 0.841 \quad \neq \quad \sin^3(1) \approx (0.841)^3 \approx 0.596$$

0.9 Probability: Arrangements, Permutations, Combinations, Factorials

One of the great applications of mathematics deals with our addiction or aversion to risk, and our attempts to predict the unpredictable. People addicted to risk can't stay away from the casinos or are always playing games of chance. People who are risk averse load themselves with insurance. (Those who take no risk own casinos or insurance companies!)

When you take risks you are always calculating probabilities, whether or not you think you are using math. Calculating probabilities often ultimately involves Calculus but we must start with simple math or algebra.

Probability starts with a simple event. What is the probability of correctly guessing a card in a deck of 52, say the **4♥**? When there are 52 cards the probability is 1 out of 52 or 1/52. A general rule is: the probability of an event is the inverse of the number of equally likely possibilities.

Probability gets interesting and complicated when we consider several consecutive events. What is the probability of guessing two cards in a row – say the **4♥** and the **9♣**? Two conditions must be specified before answering this compound problem.

1: Arrangement: Does the order or the precise arrangement matter?
2: Replacement: Is the card replaced after each guess?

The order that teammates are chosen does not matter because the team is identical whether it consists of A: Jack and Jill or, B: Jill and Jack. In probability lingo a **combination** is a team where order does not matter.

The order does matter for prizes. The scenario is different if A: Jack won first prize (fell down and broke his crown) and Jill won second or, B: Jill won first prize, and Jack won second. In probability lingo a **permutation** is a team picked in a precise order.

Any team with more than 1 member has more permutations than combinations.

Replacement occurs when a (six-sided) die is thrown or a (two-sided) coin is flipped more than once. You can get ten sixes in a row or ten heads in a row. Replacement does not occur when choosing a team because you cannot choose the same person twice. When you are dealt a hand of cards, even if you draw new cards, the cards you discard are not replaced in the deck until the hand is over. If each card were recorded and replaced, you could potentially get a Poker hand with 5 Aces of Spades. Allowing replacement increases the number of possibilities.

Perhaps the most profound and fundamental finding of probability theory is that

The number of possibilities of a string of events is the product of the number of possibilities of each individual event.

Let's return to the number of possibilities and the probability in our card guessing game. After you guess the **4♥** and remove it from the deck there are only 51 cards left so the probability of guessing the **9♣** is 1/51. Using the fundamental concept in probability – that the number of events or the probability of a string of events involves multiplying the number of events or the probabilities of each of the individual events – the number of ways we can pick two specific cards in a specified order is 52 × 51 = 2652 and the probability of doing that is 1/2652. If you see anyone doing it correctly as part of a magic trick, it is a trick and they are cheating.

The number of specific arrangements or permutations of all 52 cards is 52 factorial (52!) or 52×51×50×…×3×2×1 ≈ $8.0658(10)^{67}$, an inconceivably huge number. It is the astronomical number of permutations or combinations that make the universe and all that is in it so very diverse, so very interesting, and so very possible.

Figure 0-30 illustrates that the factorial, *n*! 1: is the number of unique arrangements or permutations and, 2: starts small but increases faster and faster as *n* increases. If, for example, there is only one colored ball there is 1 = 1! permutation (the lower left yellow box). When there are 2 differently colored balls, the number of permutations increases to 2 = 2! (the middle yellow box), for example, turquoise first and purple second or vice-versa. When there are 3, the number of permutations increases to 3! = 6 (the furthest right yellow box). When there are 4, the number of permutations increases to 4! = 24. I didn't illustrate the case of 5 colors because that involves 5! = 120 permutations.

Note that while there are 24 *permutations* or unique arrangements of the four colored balls, if you don't care about the order in which the balls were picked, there is only one *combination* (or team) of the four colored balls.

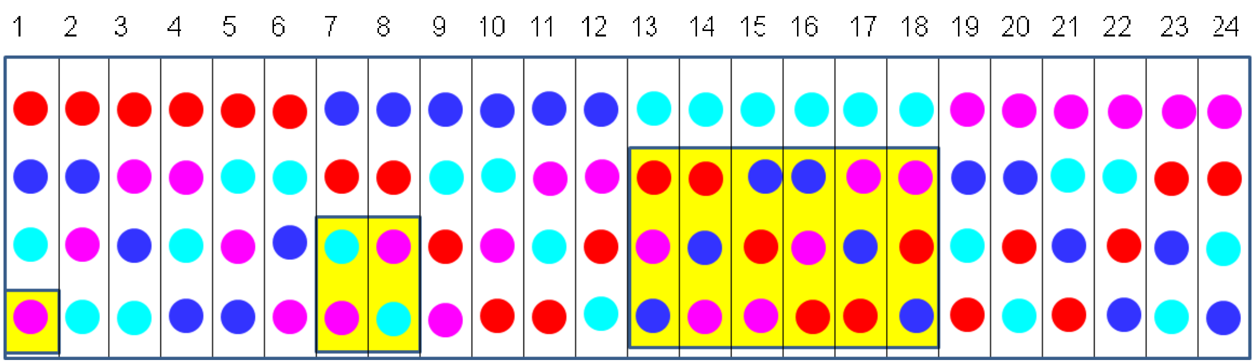


Figure 0-30 Illustration that the number of unique arrangements or permutations of 1, 2, 3, and 4 distinctly colored balls when no replacement is allowed is equal to *n* factorial (*n*!).

Now we are ready for the equations for the number of permutations and the number of combinations. If a game consists of picking *k* items (e. g., 5 cards) from a basket of *n* items (e. g., a deck of 52 cards) when none is replaced, then the number of unique arrangements or permutations of is written as ${}_nP_k$ and is given by Eqn. 0-37.

$$\boxed{{}_nP_k = \frac{n!}{(n-k)!}} \qquad \text{Eqn. 0-37}$$

Remember that **if** the order of choosing is not distinguished then there are fewer distinct teams or combinations than distinct arrangements of prize winners or permutations. If a game consists of picking *k* items (e. g., 5 cards) from a basket of *n* items (e. g., a deck of 52 cards) when none is replaced, then the number of distinct teams or hands or combinations is written as ${}_nC_k$. It is equal to the number of permutations divided by *k*!, and is given by Eqn. 0-38.

$$\boxed{{}_nC_k = \frac{{}_nP_k}{k!} = \frac{n!}{k!(n-k)!} \equiv \binom{n}{k}} \qquad \text{Eqn. 0-38}$$

You will see Eqn. 0-37 and Eqn. 0-38 again in Chapter 1. The expression, $\binom{n}{k}$ in Eqn. 0-38 is standard shorthand for all these factorials.

Problem: How many unique ways are there to pick 5 cards from a deck?
Solution: This is the number of permutations of 5 objects drawn from a deck of 52.

$${}_{52}P_5 = \frac{52!}{(52-5)!} = 52\times 51\times 50\times 49\times 48 = 311{,}875{,}200$$

There are almost 312 million distinct permutations of the 52 cards.

Problem: How many different 5-card hands are there?
Solution: Remember that a 5-card hand is a team. So, all 5-card hands with the same cards are the same no matter what the order in which they were picked. So, what we want is the number of combinations of 5 objects drawn from a deck of 52. Note that any 5 card hand can be drawn in 5! = 5×4×3×2×1 = 120 ways. Therefore, the number of distinct Poker hands (i. e., combinations) is only

$${}_{52}C_5 = \frac{{}_{52}P_5}{5!} = \frac{52!}{5!(52-5)!} = \binom{52}{5} = \frac{52\times 51\times 50\times 49\times 48}{5\times 4\times 3\times 2\times 1} = 2{,}598{,}960$$

Problem: A: How many different 5-card Poker hands are there with 4 Aces? B: What is the probability of getting a hand with 4 Aces?
Solution: The only difference between hands of 4 Aces is in the 5^{th} card. After choosing the 4 aces, there are 48 cards left, so that there are 48 distinct hands with 4 Aces. The probability of getting such a hand (without discarding and drawing other cards) is therefore 48/2598960 = 1/54145 This means that if you were a Poker addict who played 150 hands of 5-card Poker a day, you would be dealt 4 Aces about once a year.

If a game consists of picking k items with replacement and n possible choices, the number of unique arrangements is n^k. (When replacement occurs we use the term arrangement but not permutation.)

Replacement occurs (as I said) when tossing coins or throwing dice.

	Toss #1	Toss #2	Toss #3
Game #1	H	H	H
Game #2	H	H	T
Game #3	H	T	H
Game #4	H	T	T
Game #5	T	H	H
Game #6	T	H	T
Game #7	T	T	H
Game #8	T	T	T

Table 0-5 There are 8 arrangements of 3 coin flips.

Case #1: Toss a coin three times. Then n = 2 (sides of a coin) and k = 3 tosses, so there are 2^3 = 8 arrangements of **H**eads and **T**ails, shown in Table 0-5. Only one arrangement each produces 3 Heads or 3 Tails and 3 arrangements each produce 1 Head or 2 Heads.

Case #2: Throw a die 10 times. Then $n = 6$ faces on a die and $k = 10$ throws, so there are exactly 6^{10} or about 60 million arrangements. If we want the unique event of throwing the 6 ten times in a row then the probability is $1/6^{10}$ ($\approx$ 1 in 60 million)!

Replacement also exists in the more vital game of constructing the molecules of life's genetic codes. Figure 0-31 illustrates a tiny segment of DNA. It contains strings of 4 different molecules abbreviated as **A**, **C**, **G**, and **T** that stick out like hooks to join the two strands of the double helix. An **A** can only bond to a **T** and a **C** can only bond to a **G**.

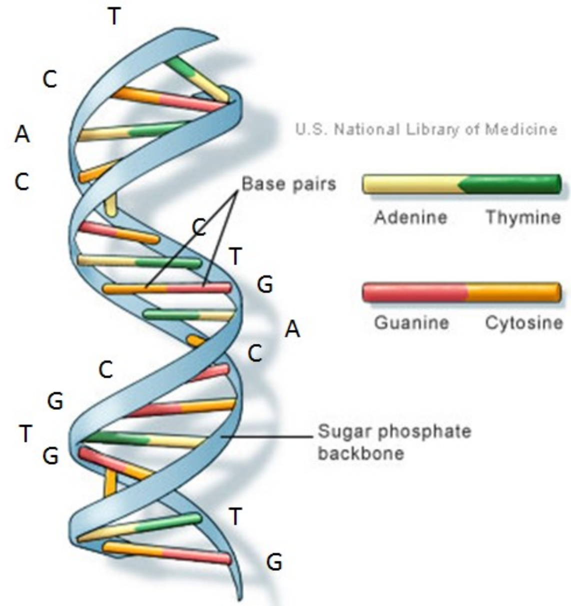


Figure 0-31 Segment of a DNA molecule.

Each strand of a single human chromosome contains roughly 150 million of these hooks or nucleotides. When DNA is replicated, so that life can continue to the next generation, every three nucleotides form a unit much like an electric plug. Since each has four possible molecules, each triplet has $4\times4\times4 = 4^3 = 64$ arrangements. Considering that there are roughly 50 million triplets per chromosome (and we have 23 chromosomes) no wonder no two people (other than identical twins) are or ever will be exactly alike without genetic engineering, even given that long segments of each chromosome are identical for all of us.

Probability vs. Statistics

Probability and Statistics are related fields of mathematics but there are differences. When you think of crossing the street you mentally calculate the probability of getting across safely. After you have either crossed or not made it across the street, you have become a statistic. Statistics involves analyzing data, such as the impact of smoking on longevity or the success of a medical procedure such as cardiac bypass or of a particular batter against a particular pitcher to help predict probabilities of future outcomes. We will return to probability when treating series in Chapter 1 and will return to probability and statistics at the end of Chapter 6.

A GENTLE PARTING WORD

If you have mastered the material in this chapter there is a 100% probability you are ready for Calculus. If you really work hard at the Calculus you will actually feel your I. Q. rising. That will be your greatest reward. You won't need compliments from anyone.

Sage Advice: If you have not fully mastered the material in this chapter but are itching to go ahead that's OK. Just be sure that any and every time you have trouble because you forgot some BC (Before Calculus) math, be sure to look back. That way you will kill two birds with one stone – you'll learn Calculus better and master the BC math sooner.

A final thought for this chapter! Tom Sawyer got the kids to whitewash the fence for him. They even begged and paid him to do it – all because he told them he was having fun and made it enticing. Mark Twain pointed out that Tom learned that day that the difference between work and play is that work is something you have to do and play is something you choose to do.

> Tom said to himself that it was not such a hollow world, after all. He had discovered a great law of human action, without knowing it – namely, that in order to make a man or a boy covet a thing, it is only necessary to make the thing difficult to attain. If he had been a great and wise philosopher, like the writer of this book, he would now have comprehended that Work consists of whatever a body is obliged to do, and that Play consists of whatever a body is not obliged to do.

So, get to ~~work~~ play! I made this book witty (well, as witty as I can) and kept it short for you. I've done everything I could for you. You must do the rest. You know it won't be easy, but,

After all is said and done
It may turn out to be fun,
And then we will all have won.

And remember, I will be with you and rooting for you all along Your Royal Road to Genius!

CHAPTER 1: THE BIRTH OF CALCULUS
AS EASY AS π, OR THE METHOD OF EXHAUSTION
RATES, SUMS, LIMITS, DERIVATIVES AND INTEGRALS

Two of the extraordinary things that Calculus does are to find 1: totals or sums (including lengths, areas, and volumes) and, 2: rates of change of processes or phenomena that are non-linear (i. e., curved).

1. Totals or sums are called integrals in Calculus.
2. Rates of change (the basis of forecasts) are called derivatives in Calculus.

Calculus also shows that totals and rates are inverse problems, much like multiplying *and dividing are inverses.*

The roots of Calculus were established 2000 years before Calculus was finally invented. This incredibly long gestation period is the strongest evidence of how astoundingly great an intellectual accomplishment Calculus is, and why it typically proves so difficult to master. If it were easy it would have been formulated much, much earlier in human history. So many other great and extraordinary accomplishments came before Calculus in good part because they were easier.

This chapter explores the roots of Calculus, proceeding right up to the integral and the derivative. So, how exactly did the deep roots of Calculus get started?

1.1 Infinite Series
Back to the Past: The First World Series

Blame it all on Zeno of Elea, who lived shortly before Socrates. In his time philosophy was the rage because they didn't have video games. One of the burning issues was, "Does Nature consist of the One or the Many?" To prove it consists of the One and to silence the opposition, Zeno proposed some paradoxes. In one of these, he proved beyond all doubt that Achilles could never beat a tortoise in a race. Simply give the tortoise a lead, and then start the race. By the time Achilles reaches the tortoise's starting point, the tortoise has moved ahead to his second position. By the time Achilles reaches the tortoise's second position the tortoise has moved on to his third position and so on, ad infinitum. Thus, it is plain that Achilles will never reach the tortoise.

Baloney, you say! But can you find the Achilles' heel in the argument? It is the assumption that an infinite series has an infinite sum. But we all know at least one infinite series that has a finite sum, namely,

$$\boxed{S_n = \sum_{i=0}^{n}\left[\frac{1}{2}\right]^i \equiv 1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots..\frac{1}{2^n} = 2-\frac{1}{2^n}}$$

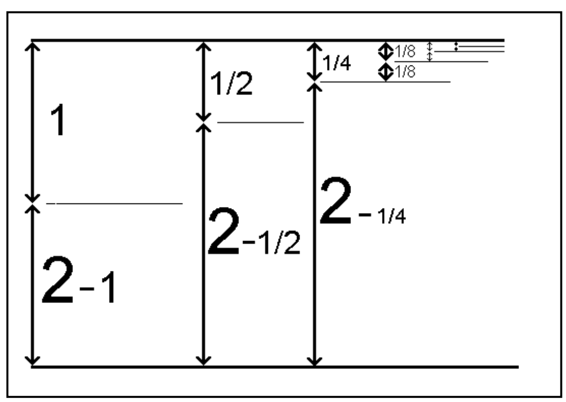


Figure 1-1 Successive sums of the series, 1+ ½+¼+… and their approach to 2.

The sum of this series, $S_n \to 2$ as $n \to \infty$. Figure 1-1 shows that and also shows that the difference between the sum and 2 is equal to the last term. Thus, 1+½ = 2-½, 1+½+¼ = 2-¼, 1+½+¼+⅛ = 2-⅛, etc.

This is an example of a....

Geometric Series,

A geometric series is one where each succeeding term is larger (or smaller) than the previous term by a constant ratio, *r* or *x*. Here, I use *x* because it is the typical variable, and the constant, *a* is the first term.

$$S_n = \sum_{i=0}^{n} ax^i \equiv a + ax + ax^2 + ... + ax^n \qquad \text{Eqn. 1-1}$$

To find the sum of a geometric series requires a step of genius. Multiply each term of the series by *x*. This yields a second series whose sum is $x \cdot S_n$,

$$xS_n = \sum_{i=0}^{n} ax^{i+1} \equiv ax + ax^2 + ax^3 + ... + ax^{n+1}$$

This was the step of genius because its hidden motivation is to subtract the second series from the first. Then, every term on the right hand sides cancels except for the first term in the first series and the last term in the second series. This gives,

$$(1-x)S_n = a - ax^{n+1} = a(1-x^{n+1})$$

To solve for S_n, simply divide both sides by $(1 - x)$. The result is the general equation for the sum of a geometric series,

$$\boxed{S_n = \frac{a(1-x^{n+1})}{(1-x)}} \qquad \text{Eqn. 1-2}$$

With Eqn. 1-2 you can calculate total savings after saving \$*a* per month for *n* months. Set $x = (1 + IR/12)$ where *IR* is the annual interest rate. But for now, lets use the series of Fig. 1-1, with initial term, $a = 1$, ratio, $x = ½$ and $n = 5$ to show that Eqn. 1-2 works.

$$S_5 = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} = \frac{63}{32} = \frac{1(1-0.5^6)}{1-0.5}$$

Wow! Eqn. 1-2 does work! Thank God someone else derived it for us! Eqn. 1-2 may have been the first hint of an astounding revelation, namely that a long or even infinite series of numbers can equal a simple expression that may have a finite value. You probably won't believe how frequently series are used so I won't bother telling you….

OK, since you insist, I will give you a few advance hints of the power of series. But first, let's introduce…

Arithmetic Series

An arithmetic series is one for which each succeeding term increases (or decreases) by a constant amount, such as 1, 2, 3, 4…. There is a famous story involving arithmetic series and the math genius, Karl Friedrich Gauss. You decide whether or not it's true. When Gauss was about 9 years old his teacher decided to punish the class. "Add all the integers from 1 to 100." Gauss almost immediately wrote the correct answer on his chalkboard (paper was in short supply and they didn't have computers in those days). He used a trick. He noted that the first term plus the last term (1 + 100) adds up to 101. So does the 2nd term plus the next to last term (2 + 99 = 101). There are 50 of these equal pairs. So instead of adding, he multiplied 101×50 = 5050. Multiplying is harder than addition, but **can be much, much faster.**

The general formula for the sum of an arithmetic series is found by adding it to the exact same series with all terms written in reverse order. If each series contains *n* terms then the two series together contain *n* equivalent pairs (linked by the arrows). If the first term of the series is *a* and the difference between successive terms is *a*, then the sum of each pair is, *a*(*n*+1), and the result is Eqn. 1-2.

$$\begin{array}{l} S_n = \sum_{i=0}^{n} ia \equiv a \quad + \quad 2a \quad + \quad 3a + \ldots \quad + na \\ \qquad\qquad\quad \updownarrow \qquad\quad \updownarrow \qquad\qquad \updownarrow \qquad\qquad \updownarrow \\ S_n = \sum_{i=0}^{n} ia \equiv na + (n-1)a + (n-2)a + \ldots + a \end{array}$$

Eqn. 1-3

The double arrows link identical pairs of numbers. Adding these *n* identical pairs equals 2*S*. Then, dividing by 2 yields Eqn. 1-4, the formula for the sum of an arithmetic series.

$$\boxed{S_n = \frac{an(n+1)}{2}}$$

Eqn. 1-4

Using Series to Evaluate Transcendental Functions

Series are used to approximate many transcendental (i. e., impossible) functions. The series are built into calculators and computer software and hardware. An example is the sine function, which is given by the Taylor Series (see Chapter 4) of Eqn. 1-5.

$$\boxed{\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots.}$$

Eqn. 1-5

On a historical note, this series and infinite series for other trig functions were apparently worked out by the amazing Indian Mathematician, Madhava shortly after the year 1300!

Problem: Use Eqn. 1-5 to approximate sin(30°), which we know equals 0.5 exactly. **Solution**: First, convert 30° to radians. 30° ≈ 30/57.3 ≈ 0.524. Then substitute

$$\sin(0.524) \approx 0.524 - \frac{0.524^3}{3!} + \frac{0.524^5}{5!} - \frac{0.524^7}{7!} \approx 0.5003$$

Since the exact value is 0.5, for this case, the value provided by the first 4 terms of the series is, as they say, "Good enough for Government work!"

Evaluating π with the Taylor Series for the Inverse Tangent

Remember that inverse tangents and inverse sines are angles. So is π. When Newton developed Calculus he found a series for the inverse sine and realized he could use it to approximate π to zillions of figures. He spent so much time that he confessed, "I am ashamed to tell you to how many figures I carried these computations, having no other business at the time". Sounds more like bragging to me!

We use the Taylor Series for the inverse tangent instead to evaluate π.

$$\boxed{\tan^{-1}(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots} \qquad \text{Eqn. 1-6}$$

To get π from Eqn. 1-6, set x = 1 and multiply by 4 because the angle whose tangent is 1, i. e., $\tan^{-1}(1) = \pi/4$.

$$4\tan^{-1}(1) = \pi = 4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots\right)$$

The only trouble with this series is that you need to calculate many terms unless you are smart enough to graph individual sums (the red line in Figure 1-2) and observe that while they oscillate up and down, the average of two consecutive sums (the blue line in Figure 1-2) is much steadier and should produce a pretty accurate result. For example, the average of a series that extends to 1/7 and one that extends to 1/9 (so that you add half of 1/9) is,

$$\pi \approx 4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{2} \times \frac{1}{9}\right) \approx 3.117$$

The error of this estimate for π is only 0.77%.

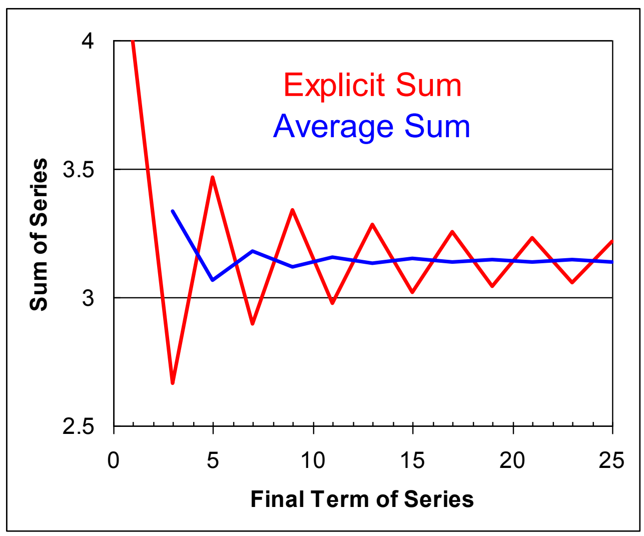


Figure 1-2 Approximating π with the sum of series for the inverse tangent of 1. Red curve shows the individual sums; blue curve shows the averages of two consecutive sums.

1.2 Approximation, Estimation, and Errors: Math Ain't Always Perfect

The series for the sine and for the inverse tangent give almost perfectly accurate results if we include an infinite number of terms. But that would take an infinite amount of time. So, infinite series force us to 1: relinquish the idea of perfection and, 2: decide how accurate we need to be and therefore how many terms to calculate.

Most of us are so accustomed to thinking math is exact that we are stunned to encounter any math problem that does not have an exact answer or solution. Consider statistics, which is a mathematical confession of ignorance and might be called the math of errors. Both Nuclear and Human behavior are so complex that there is no universal law to predict with certainty outcomes of events such as electrons or elections with certainty. So, for elections, we survey a bunch of people and assume (i. e., hope) they represent an unbiased sample. But the next bunch of people or the same bunch the next week will almost surely give a somewhat different result, so we are forced to face the fact that some problems have built in uncertainty.

When it comes to money, we easily relinquish perfect accuracy. Consider that you earn \$10.00 per hour and work for 62 minutes. You should earn \$10.333333…. Your boss will probably try to get away with paying you only \$10.00, but you might demand the extra. However, he cannot pay you the exact extra \$0.333333… because the smallest amount of change in the USA is a penny or \$0.01. So if your boss gives into your demand you will get an extra 33¢ or, with an extremely generous boss, 34¢.

For the next concession to mathematical imperfection, try to calculate total annual earnings for the 3 cases of weekly salaries in Figure 1-3 (the turquoise areas).

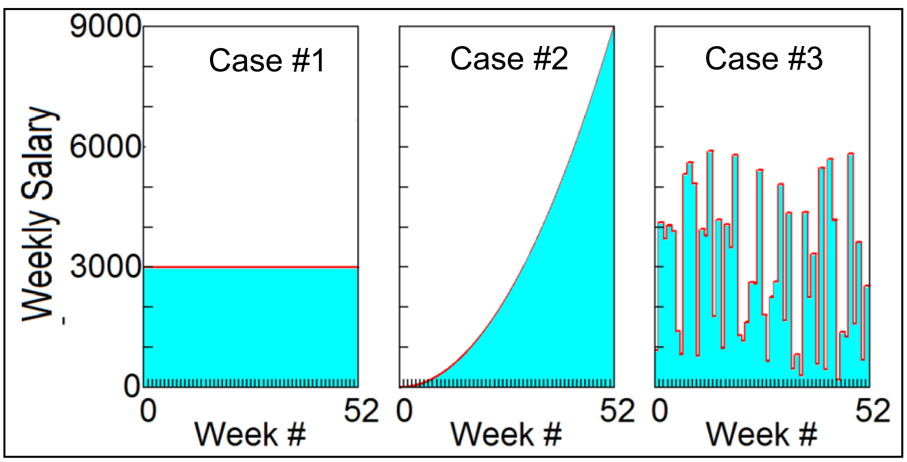


Figure 1-3. Three cases of weekly salary. Case #1 is constant, Case #2 is a parabola and Case #3 is random. Use these to calculate or estimate annual earnings.

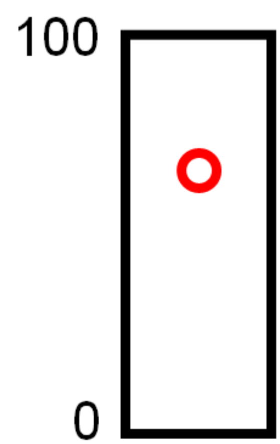


Figure 1-4. It's not possible to find the height of the red circle exactly!

Case #1 is easy because it is the area of a rectangle. If you earn \$3000 per week for 52 weeks (assuming it is exactly equal to 1 year) then simply multiply to get \$156,000. Case #2 is tougher and is a job for Calculus. The weekly salary is a parabola that starts at \$0 and ends at \$9000. In a few pages you will find that the average height of a parabola starting at 0 is ⅓ the peak height, Therefore, the average salary is ⅓ the peak salary or \$3000. So Case #2 has the same area as Case #1, namely \$156,000.

But what about Case #3? Calculus won't help. This is not a job for a mathematician – it is a job for Alexander the Great. According to legend, anyone who undid the Gordian knot would conquer the world. Alexander didn't bother with fine details. He simply hacked the knot with his sword. And that is what we have to do here. Simply look at the random peaks and valleys and estimate the average using your eyeballs. You will see that the weekly salary averages out to about $3000.

Another example of imperfection is to measure or estimate the height (from 0 to 100) of the **o** in Figure 1-4. You will always make some error, no matter how tiny. If you can't measure or observe with perfect accuracy it will be impossible to predict exactly when the **o** will hit the bottom if it falls. Remember that a chain is only as strong as its weakest link. So if the measurements are imperfect, don't worry if the math is also imperfect so long as the math errors are no larger than the measurement errors.

Now that you are finally ready to accept math that is not perfectly accurate, we are ready to embark on the incredible journey that will lead us to Calculus. We start with…

1.3 Zeno's Paradoxes and the Method of Exhaustion

Zeno's paradoxes proved to be great favorites of lawyers and philosophers, who have used them to waste centuries of labor and destroy countless lives. However, some value did come of Zeno's paradoxes. Eudoxus (who knew Plato and Aristotle but never could get them to learn much math) gave Zeno's paradoxes a mathematical twist in order to find the area under a parabola.

What's so tough about finding the area under a parabola? Well, try finding the area under or in any curve. Sure, you know that the area of a circle is πr^2, but the only reason you know it is that you were told. In fact, that is the only reason you know the value of π. Finding the area under or inside curves is one of the great uses of Calculus. No one needs Calculus to find the area of a rectangle - even a baby can do that - and the area of a triangle is only a little trickier, but finding the area of circles or other curves such as parabolas requires ingenuity. In Section 0.7 we were able to find the slope of the parabola using only algebra because we have the quadratic formula, but algebra alone will not give an analytic expression for the area under the parabola. And that leaves us virtually no hope to find equations for the area under other, more complicated curves.

What's so important about areas (alias integrals) anyway? Areas represent quantities such as total distance traversed or total earnings over an extended time period, and you need Calculus to calculate these when the rates are not constant. In Calculus, the area under or in a curve is called the integral.

Here is how Eudoxus found the area under a parabola. His step of genius began with relinquishing the idea of perfection. Since he could calculate the area of triangles, he could find the approximate area under a parabola (or any curve) by squeezing as many successively smaller triangles as possible either under or over the curve and adding the area of the triangles.

Take the section from $x = 0$ to $x = 1$ of the simplest parabola, $y = x^2$. You should see that the area under the parabola in Figure 1-5 is less than 1, which is the area of the square that extends from (0, 0) to (1, 1). Eudoxus then began to chip away at the square. He drew a diagonal straight line connecting the endpoints of the parabola. This divides the square into two right triangles, each with area ½. He discarded the upper left, yellow triangle because it lies entirely above the parabola. The remaining area is ½. It consists of the turquoise area under the parabola and the excess, purple area. Eudoxus then drew a second yellow triangle to discard because it also lies above the parabola and joins it at 3 points, (0, 0), (½,¼) and (1, 1). The altitude of this triangle is ½ -¼ = ¼ and its base is 1, so its area is ⅛. The remaining areas now total ⅜.

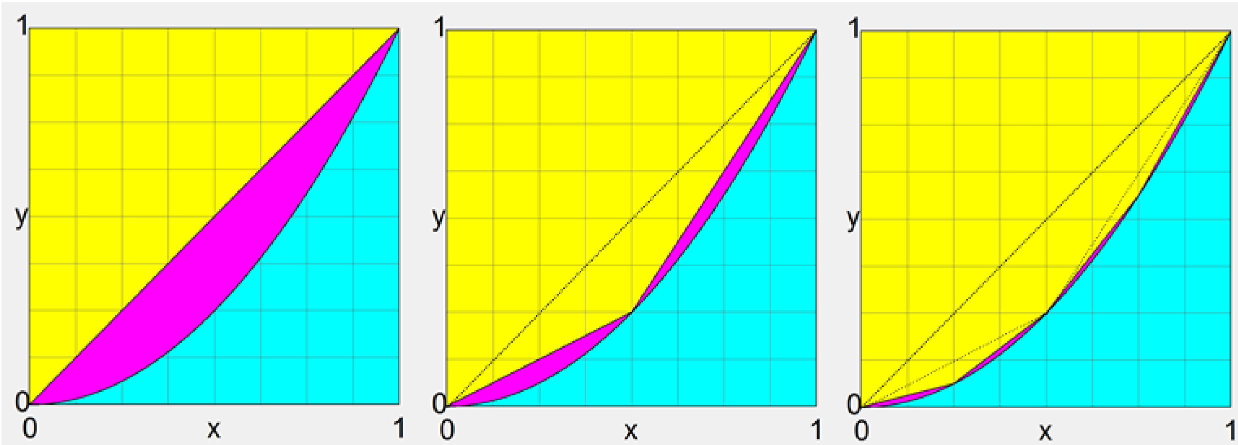


Figure 1-5 Approaching the parabola by slicing off triangles. The residual purple area keeps getting smaller and eventually vanishes. This is the Method of Exhaustion.

At the next step, discard two more triangles that partly fill in the two slivers between the previous triangle and the parabola, touching the parabola at x = ¼ and x = ¾ respectively. The area of these two triangles sums to $^1/_{32}$. Now, the purple and turquoise areas equal $^{12}/_{32}$ - $^1/_{32}$ = $^{11}/_{32}$. We can continue this process indefinitely, each time cutting down the purple area as we get closer to the parabola. Each succeeding step doubles the number of new triangles but adds only ¼th the area of the previous triangles. This leads to an infinite (geometric) series of triangles that ultimately merges with the parabola. Thus the area under the parabola = ⅓, the sum of the geometric series,

$$A = 1 - \frac{1}{2}\left(1 + 0.25 + 0.25^2 + \ldots\right) = 1 - \frac{1}{2}\left(\frac{1}{1 - .25}\right) = 1 - \frac{2}{3} = \frac{1}{3}$$

You can probably see why this is called the Method of Exhaustion. Believe it or not, Eudoxus was at the doorstep of Calculus, yet it took another 2000 years to cross it!

Archimedes, perhaps the greatest mathematician, physicist, and nudist of the Ancient World, used the Method of Exhaustion to find areas and volumes of other curves and solids. We will now use it to find the area of a circle and the value of π. By 2000 BC, both the Egyptians (3.16) and Mesopotamians (3.12) knew that π, the ratio of the circumference of a circle to its diameter is about 3.14. How did they know this? They probably measured the circumference and diameter with a string. Don't knock what works. Try it yourself! Get a tape measure and for starters, measure a friend's waist line just to show how friendly math can be. (You will find that it is not so easy to measure the diameter but I'm sure you will work it out.) Do this for a circular object as well. Then take the ratio of circumference to diameter and see how close you get - to π that is! Of course, measuring is only a starting point because it always involves some errors. The Method of Exhaustion reduces these errors.

Squaring, Octagoning and Polygoning the Circle

We will now follow in Archimedes' footsteps. Begin by drawing a circle with radius, $r = 1$. Next, inscribe one square inside the circle and circumscribe a second square outside the circle, as in Figure 1-6. Find the area and perimeter of each square. Each side of the circumscribed square is equal to the diameter or, $2r = 2$, so its perimeter is $8r = 8$ and its area is $(2r)^2 = 4$. For the inscribed square the Pythagorean Theorem shows a side is $r\sqrt{2} = \sqrt{2}$. Its perimeter is then $4\sqrt{2} \approx 5.66$ and its area is 2.

The average perimeter of the two squares is 6.83, or 8.7% larger than 2π, the circle's circumference. The average area is 3, or 4.5% smaller than π, the area of the circle. The errors are not bad considering how crude it is to approximate a circle by squares!

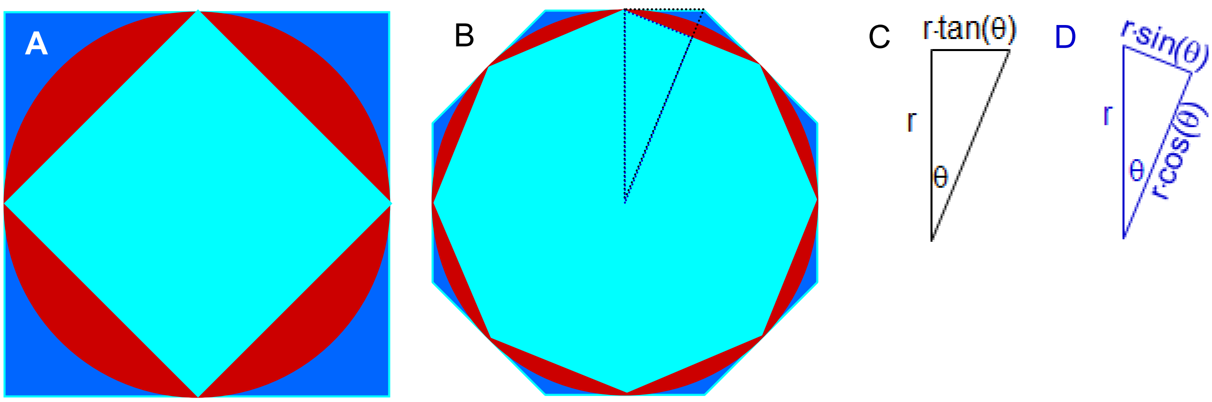


Figure 1-6 Approximating the circle with circumscribed and inscribed squares (A) and octagons (B). Right triangles for circumscribed (C) and inscribed (D) polygons are shown in the right two diagrams.

Let's get more accurate! Inscribe and circumscribe octagons. The octagons embrace the circle more closely than squares. Both the external blue area of the large octagon and the exposed red area of the circle are much smaller than when squares were used.

Now for the math! Divide each octagon into 16 right triangles to use Pythagoras. Each right triangle has central angle, $\theta = 360°/16 = 22.5°$. The perimeter and area of each octagon is 16 times the base and area of each of its triangles. Figure 1-6C,D show that the circumscribed octagon's perimeter $16r\cdot\tan(22.5°) \approx 6.63$ and area, $A = 16\cdot\frac{1}{2}r^2\cdot\tan(22.5°) \approx 3.31$ while the inscribed octagon has perimeter, $16r\cdot\sin(22.5°) = 6.12$ and area, $A = 8r\cdot\sin(22.5°)\cdot r\cdot\cos(22.5°) \approx 2.83$. The average perimeter of the octagons is 6.37, or only 1.45% larger than the perimeter of the circle. The average area is 3.07 or 2.24% smaller than the area of the circle. Approximating the circle with octagons (instead of squares) reduces the average errors for estimating π to less than 1.9%.

This procedure employs trigonometry. That makes it easy to generalize to a polygon with any number of sides. If the polygon has n sides, there are $2n$ right triangles, and each has a central angle, $\theta = 2\pi/(2n)$ radians. Figure 1-6 C and D show that the bases and altitudes of the inscribed and circumscribed right triangles are given by Eqn. 1-7.

We find the values of the sines, cosines and tangents from a calculator. Archimedes didn't have a calculator so he slaved away using half angle formulas such as Eqn. 0-13b from Section 0.4, namely, $\sin(A/2) = [\frac{1}{2}(1 - \cos(A))]^{0.5}$. In this way, he worked his way to a polygon with 64 sides. His estimates for π were very good, low by 0.04% for area and high by 0.02% for perimeter, for an average error of 0.01%!

Triangle	Base	Altitude	Area
Inscribed	$r\sin(\theta)$	$r\cos(\theta)$	$\frac{1}{2}r^2\sin(\theta)\cos(\theta)$
Circumscribed	$r\tan(\theta)$	r	$\frac{1}{2}r^2\tan(\theta)$

$$\boxed{\theta = \theta_n = \frac{2\pi}{2n}} \qquad \text{Eqn. 1-7}$$

These examples have not exhausted the possibilities of the Method of Exhaustion, but they probably have exhausted you. The moral of the story is, when anyone tells you that something is as easy as π, beware! But as we have seen, mathematical series that are derived from Calculus <u>are</u> as easy as π.

Using Blocks to Find the Area under Curves: The Riemann Sum or Quadrature

For 1800 years after Archimedes, no one came any closer to creating Calculus. One reason the Greeks failed to develop Calculus is that they didn't know their algebra. Algebra was developed by the Hindus and Moslems (who named it), but neither advanced a single step closer to Calculus. Progress came only in the 1500's, after the Europeans finally accepted the infidel's algebra and used it to improve upon the Classical texts of mathematics and physics they had recently rediscovered.

Algebra makes it easy to approximate the area under curves by fitting them to a bunch of skinny rectangles or blocks and then calculating the area of the rectangles, which is easy to do. Why, you might almost say that algebra makes it almost like playing with blocks! This, of course, is extremely important for mathematical blockheads. By the way, the person who developed this approximate technique was no blockhead, but Bernhard Riemann, a brilliant mathematician. That is why the procedure is called a **Riemann Sum**. It's also called **quadrature**.

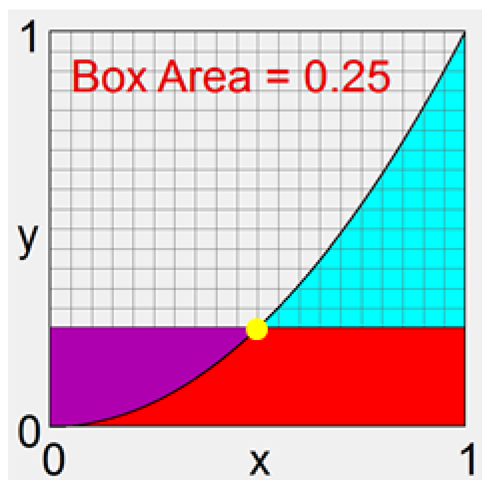


Figure 1-7 Approximating the parabola with a rectangle.

Start with a single fat rectangle to approximate the parabola, $y = x^2$ from $x = 0$ to $x = 1$, as in Figure 1-7. The rectangle has two colors – purple for the part above the parabola and red for the part below. The purple area is excess area that should not be included. The turquoise area (below the parabola and outside the rectangle) is the area that should have been included but was not.

The problem with rectangles is how to choose their **height**. The best choice is the one where the excess purple area exactly equals the deficit turquoise area. But that assumes we know the solution to the problem we are trying to solve, and we don't know it.

The heights of the rectangles are set by the yellow points or the values of x where the rectangle tops touch the parabola. Figure 1-8 shows the three standard choices.

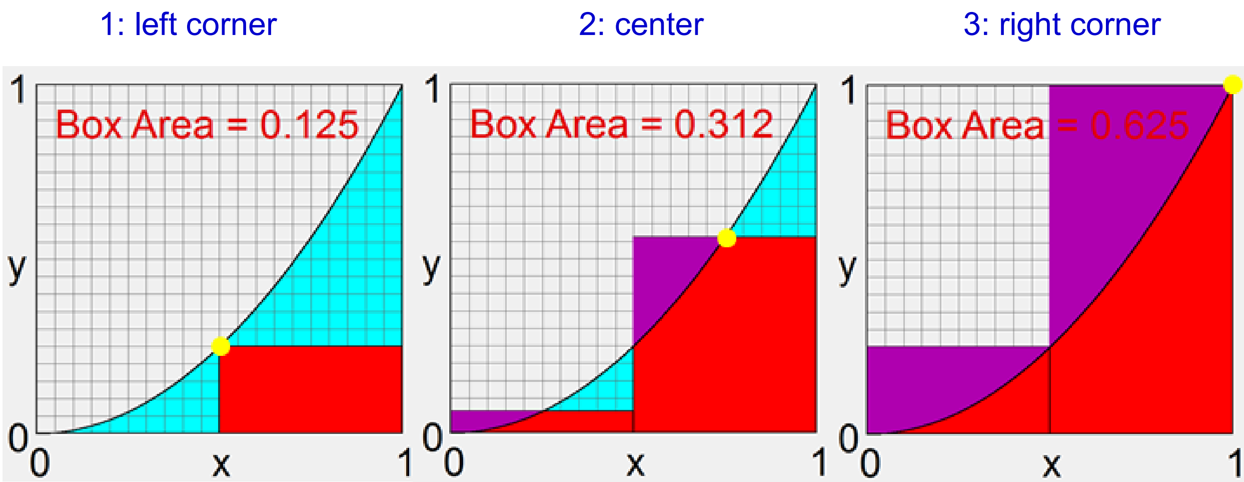


Figure 1-8. Using boxes to estimate the area under the parabola when it coincides with the yellow dots at left, center, or right side of the box tops. The center point gives the best match.

In Figure 1-7 we chose the center point (x = 0.5). At this point, $y = x^2 = 0.5^2 = 0.25$. The resulting area of the rectangle is 0.25×1 = ¼. This underestimates the parabola's area of ⅓. Figure 1-8 compares the three choices when the width of the rectangles is Δx = 0.5 and shows that the center gives the most accurate estimate of the parabola's area. As Goldilocks would say using the left side is too small, using the right side is too big, but using the center point is (almost) just right.

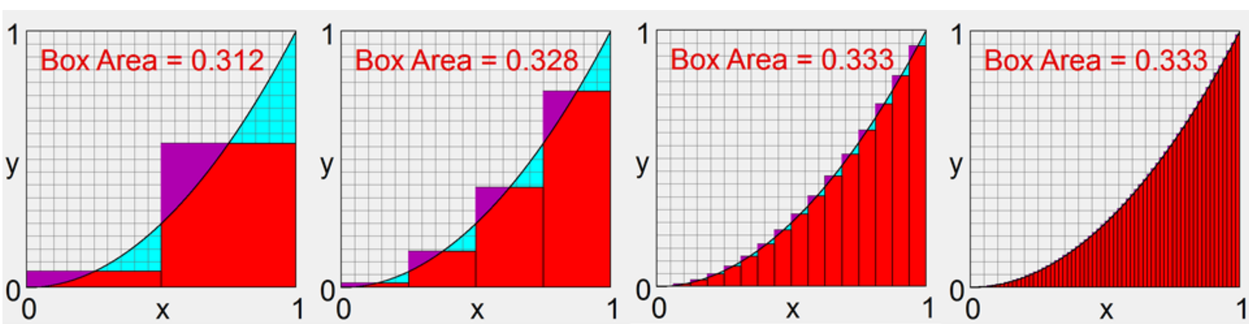


Figure 1-9 Improving accuracy of approximation to the parabola by increasing the number of rectangles from 2 to 4 to 16 to 64. For 16 rectangles A ≈ 0.33301; for 64, A ≈ 0.33331

Figure 1-9 shows that increasing the number of boxes and making each box thinner improves the fit to the parabola. When 64 boxes are used, the difference between the parabola and the boxes is almost too small to see, no matter how we choose the points where they touch the parabola.

Now, change the domain of the parabola, $y = x^2$ to $0 \le x \le b$ and find the new area. It is customary to choose n rectangles, so the width, Δx of each box is $1/n^{th}$ of b,

$$\Delta x = \frac{b}{n} \qquad \text{Eqn. 1-8}$$

To calculate the area of the i^{th} box, multiply its height, $y = x^2 = (i\cdot\Delta x)^2$ by the width, $\Delta x = b/n$. To find the total area of all boxes take the sum and substitute for $\Delta x = b/n$.

$$A = \Delta x \sum_{i=1}^{n} (i\Delta x)^2 = \left(\frac{b}{n}\right)^3 \sum_{i=1}^{n} i^2 \qquad \text{Eqn. 1-9}$$

All that is left to do is to find the general equation for the sum of squares of numbers. As you probably suspected, this requires another step of genius. So, watch the following pattern that some genius happened to notice (when the denominator is set equal to 3).

$$\frac{1^2}{1} = \frac{3}{3} \qquad \frac{1^2+2^2}{1+2} = \frac{5}{3} \qquad \frac{1^2+2^2+3^2}{1+2+3} = \frac{7}{3} \qquad \frac{1^2+2^2+3^2+4^2}{1+2+3+4} = \frac{9}{3}$$

Do you see the pattern? By induction, the general form is,

$$\boxed{\frac{\sum_{i=1}^{n} i^2}{\sum_{i=1}^{n} i} = \frac{(2n+1)}{3}} \qquad \text{Eqn. 1-10}$$

The denominator on the left hand side of Eqn. 1-10 is ½$n(n+1)$, the sum of an arithmetic series (Eqn. 1-4). Multiplying both sides of Eqn. 1-10 by Σi =½$n(n+1)$ yields,

$$\boxed{\sum_{i=1}^{n} i^2 = \left(\frac{n(n+1)(2n+1)}{6}\right)} \qquad \text{Eqn. 1-11}$$

Substituting Eqn. 1-11 into the equation for area of boxes (Eqn. 1-9) yields,

$$A = \left(\frac{b}{n}\right)^3 \left(\frac{n(n+1)(2n+1)}{6}\right) = \frac{b^3}{3}\left(1 + \frac{3}{2n} + \frac{1}{n^2}\right) \rightarrow \frac{b^3}{3} \qquad \text{Eqn. 1-12}$$

This result is remarkable. It confirms the picture that as boxes get thinner and more numerous, their total area approaches the area of the parabola because as n gets very large any term with a power of n in the denominator is so tiny it can be neglected. Eqn. 1-12 even shows the size of the error made by using boxes to find the parabola's area.

Box Method for the Area of the Cubic Curve, $y = x^3$ and for Higher Powers, $y = x^n$

To find the area under the CUBIC curve, $y = x^3$ for $0 \le x \le b$, modify Eqn. 1-9.

$$A = \Delta x \sum_{i=1}^{n} (i\Delta x)^3 = \left(\frac{b}{n}\right)^4 \sum_{i=1}^{n} i^3 \qquad \text{Eqn. 1-13}$$

Now the step of genius is to find the equation for the sum of cubes, $\sum i^3$. I looked at the pattern and made a bunch of guesses until the one below worked.

$$1^3 = \frac{(1)^2(1+1)^2}{2^2} \qquad 1^3 + 2^3 = \frac{(2)^2(2+1)^2}{2^2} \qquad 1^3+2^3+3^3 = \frac{(3)^2(3+1)^2}{2^2}$$

By induction, the general equation for the sum of cubes of integers is,

$$\boxed{\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}} \qquad \text{Eqn. 1-14}$$

Substituting Eqn. 1-14 into Eqn. 1-13 for the area of boxes for the cubic curve yields,

$$A = \left(\frac{b}{n}\right)^4 \left(\frac{n^2(n+1)^2}{4}\right) = \frac{b^4}{4}\left(1 + \frac{2}{n} + \frac{1}{n^2}\right) \rightarrow \frac{b^4}{4}$$

Once again, as n gets very large any term with a positive power of n in the denominator is so tiny it can be neglected. Then the area simplifies to ¼b^4.

Now that we have done areas for $y = x$, $y = x^2$, and $y = x^3$ (all for the domain $0 \le x \le b$) you should see a pattern by induction! Watch below, but guess before you reach the end of the line!!

$$A_{y=x^1} = \frac{b^2}{2} \rightarrow A_{y=x^2} = \frac{b^3}{3} \rightarrow A_{y=x^3} = \frac{b^4}{4} \Rightarrow \boxed{A_{y=x^n} = \frac{b^{n+1}}{n+1}} \qquad \text{Eqn. 1-15}$$

Eqn. 1-15 is a preview of a fundamental result of Calculus that the integral of x^n is $x^{(n+1)}/(n+1)$.

1.4 Limits: The Evolution from ∑ to ∫ (Sum to Integral) and from Δx to dx

Limit is a secret word I have been avoiding but whose concept I have been using throughout this Chapter. For example, the geometric series 1 + ½ + ¼ + … has a sum that approaches 2 but always falls short. Since we can make the difference as small as we like by taking more and more terms we forget the difference. In the limit as $n \rightarrow \infty$ the sum is 2.

This is also true of the Riemann sum. As n, the number of boxes increases, the width, Δx, of each box decreases. Thus, as $n \rightarrow \infty$, $\Delta x \rightarrow 0$. Of course, $n \cdot \Delta x$ remains the same but technically, we cannot allow Δx to reach 0 because if $\Delta x = 0$ then the area of each rectangle will be 0. Calculus avoids this and other similar problems by taking the limit of the area as the width, $\Delta x \rightarrow 0$ and forcing the miniscule terms to drop out completely.

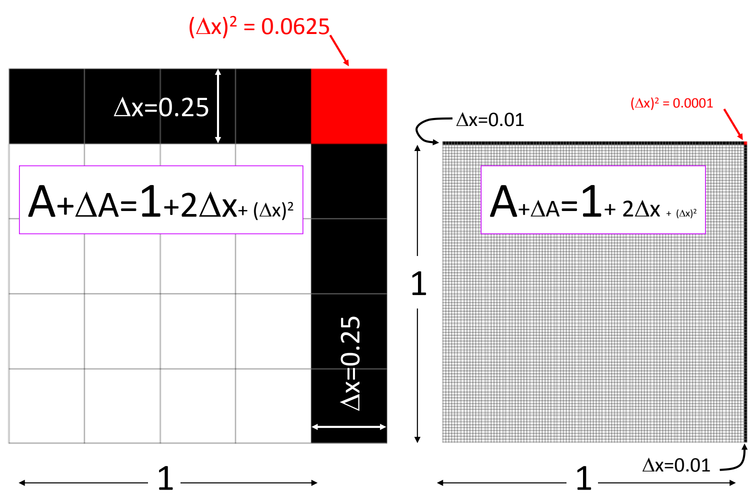


Figure 1-10 Why $(\Delta x)^2$ (red boxes) disappear.

Figure 1-10 reveals the miniscule term as the red boxes. It contains two original squares $x = 1$ on a side so the original areas are $x^2 = 1^2 = 1$. On the left, $\Delta x = 0.25$ is added to both length and width, so the new area is $(x + \Delta x)^2 = 1.25^2 = 1.5625$. The extra area consists of one red box and 8 black boxes. Neglecting the red box is like neglecting Δx^2, and it would cause us to miss 1/9th or 11.1% of the extra area. On the right, where $\Delta x = 0.01$, the extra area consists of 1 red box and 200 black boxes, so neglecting the red box would only cause us to miss 1/201st or ≈ 0.5% of the extra area.

If the original square in Figure 1-10 were 200 units on a side, increasing each side by 1 unit produces 400 black squares and still only 1 red square, which would be very difficult to see and very easy to neglect. Neglecting the red square in that case would only cause a percentage error of 1/401 ≈ 0.25%

In general, when the length, x, of the sides of a square increases by Δx, the area, A, of the square increases by,

$$\Delta A = (x + \Delta x)^2 - x^2 = 2x\Delta x + (\Delta x)^2$$

If Δx is small enough, $(\Delta x)^2 << 2x\Delta x$, so $(\Delta x)^2$ can be neglected and thus,

$$\Delta A = 2x\Delta x + (\Delta x)^2 \approx 2x\Delta x$$

Indeed, if Δx is small enough, the extra area, ΔA can be neglected. But the rate at which A increases, i. e., the **ratio**, $\Delta A/\Delta x \approx 2x$ remains finite no matter how small Δx. Taking the **limit** (abbreviated in equations as **lim**) as $\Delta x \to 0$, yields,

$$\lim_{\Delta x \to 0}\left(\frac{\Delta A}{\Delta x}\right) = \lim_{\Delta x \to 0}\left(\frac{(x + \Delta x)^2 - x^2}{\Delta x}\right) = \lim_{\Delta x \to 0}(2x + \Delta x) = 2x$$

In this beautiful expression the limiting ratio, $\Delta A/\Delta x \to 2x$ is independent of Δx and finite even though numerator and denominator each approach 0. This limiting ratio in Calculus constitutes the...

Definition of the Derivative

$$\boxed{\frac{\mathrm{d}y}{\mathrm{d}x} \equiv \lim_{\Delta x \to 0}\frac{\Delta y}{\Delta x} \equiv \lim_{\Delta x \to 0}\frac{y(x + \Delta x) - y(x)}{\Delta x}} \qquad \text{Eqn. 1-16}$$

Here we use y (in place of A) because everyone uses y. Both dy and dx are infinitesimal quantities called **differentials**. They are the limits of Δy as $\Delta y \to 0$ and Δx as $\Delta x \to 0$.

In one sense, Calculus represents the art of taking limits wisely, i. e., getting rid of tiny terms that must ultimately disappear and would make the math far more difficult. That is a key to problems involving 1: areas or sums and, 2: rates of change.

The generalizations above amount to the evolution from the sum to the integral, the difference to the differential and the ratio of differences to the derivative. Figure 1-11 shows that they resemble the evolution from ape to human. The integral and derivative are the perfect end products of evolution. Unfortunately, humans are not the perfect end product of evolution from the apes because Calculus is still difficult for most of us to master. But we're getting close!

Figure 1-11 Evolution from Ape to Human, Sum to Integral, and difference to differential.

Indefinite and Definite Integrals

Eqn. 1-12 showed that $b^3/3$ is the integral of $y = x^2$, i. e., the area under the parabola, $y = x^2$ for the domain, $0 \le x \le b$. The standard symbol for the integral is,

$$\boxed{\underset{x=0}{\overset{b}{A}}(y=x^2) = \int_0^b x^2 \mathrm{d}x = \frac{b^3}{3}} \qquad \text{Eqn. 1-17}$$

This is the **definite integral**, the integral for which the upper and lower bounds of the independent variable (in this case, x) are given and marked by the symbols at the bottom and top of the integral sign. (The terminology is similar to that used for sums.)

How would you answer if I asked you to find the area under the parabola for $2 \le x \le 5$? This is like a question on the SAT. Look at Figure 1-12. If you can't calculate the pure green area but you can calculate both the entire area and the hatched area, then the pure green area is the total area minus the hatched area. Eqn. 1-17 shows that the hatched area under $y = x^2$ is $2^3/3 = 8/3$ for $0 \le x \le 2$, and the total area is $5^3/3 = 125/3$ when x extends for $0 \le x \le 5$. The pure green area under $y = x^2$ for $2 \le x \le 5$ is then the difference, or,

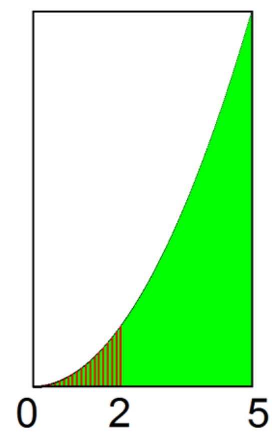


Figure 1-12 Area under $y = x^2$ for $2 \le x \le 5$.

$$\underset{x=0}{\overset{5}{A}} = \underset{x=2}{\overset{5}{A}} + \underset{x=0}{\overset{2}{A}} \rightarrow \underset{x=2}{\overset{5}{A}} = \underset{x=0}{\overset{5}{A}} - \underset{x=0}{\overset{2}{A}} \rightarrow \underset{x=2}{\overset{5}{A}}(y=x^2) = \frac{5^3}{3} - \frac{2^3}{3} = \frac{117}{3}$$

Using letters instead of numbers, the definite integral becomes

$$\boxed{\int_a^b x^2 \mathrm{d}x = \int_0^b x^2 \mathrm{d}x - \int_0^a x^2 \mathrm{d}x = \frac{b^3}{3} - \frac{a^3}{3}} \qquad \text{Eqn. 1-18}$$

If we do not state or know the limits on the independent variable, x, the integral is called an **indefinite integral**. We cannot calculate the value of an indefinite integral for the same reason we cannot calculate the area of a rectangle if we do not know its width.

Since we cannot evaluate an indefinite integral but know its functional form, we always add an unknown constant to it and write the integral without limits. For example, the indefinite integral of x^2 is $x^3/3$ plus an unknown constant, C, as in Eqn. 1-19.

$$\boxed{A(y=x^2)=\int x^2 \mathrm{d}x=\frac{x^3}{3}+C} \qquad \text{Eqn. 1-19}$$

To retrieve the original function ($y = x^2$) in Eqn. 1-19, simply take the derivative of the right hand side. The constant will disappear and we will get x^2. Indefinite integrals can indeed be beautiful general expressions, but any time you want to find the value of an integral you must specify both its lower and upper bounds.

Slopes, Rates and Derivatives

We have already discussed rates and slopes and found them for straight lines and parabolas in Chapter 0, but algebra cannot take us much further. Starting in this Chapter and continuing in Chapter 2, we will see how to approximate the slope of any curve and find exact equations for the slope of many curves.

Since we spent so much time finding the area of a parabola, let's start by finding its slope as we would with any curve - without the benefit of the quadratic formula. Once again we relinquish the hope of getting an exact answer and settle for an approximate answer.

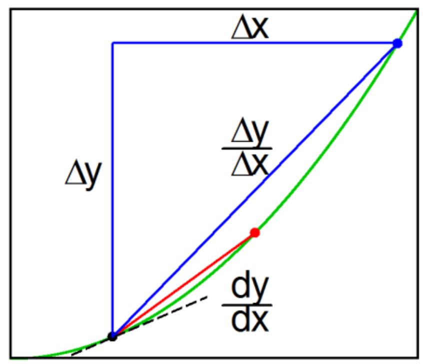


Figure 1-13 Approximating the slope of a curve.

The procedure, illustrated in Figure 1-13, is to connect two nearby points on the curve with a straight line. When the points are close, the slope of the line will be close to the slope of the curve. Fix one point and move the other closer to it. As the two points get closer the connecting line approaches the tangent to the curve. In the limit when the two points coincide, the line is tangent to the curve.

Let's do the math for the parabola, $y = x^2$, from x to $x + \Delta x$. Since we choose Δx we only need to find Δy to find the slope, or the ratio, $\Delta y/\Delta x$. To do that we first find $y + \Delta y$ at *point $x + \Delta x$, next subtract y at point x, and finally, divide by Δx.*

$$\begin{aligned} y+\Delta y &= (x+\Delta x)^2 = x^2+2x\Delta x+\Delta x^2 \\ \Delta y &= 2x\Delta x+\Delta x^2 \end{aligned} \quad \Rightarrow \quad \frac{\Delta y}{\Delta x}=2x+\Delta x \qquad \text{Eqn. 1-20}$$

The Slope of a Curve: The Derivative

Remember that we can choose Δx to be as tiny as we like. Once again, we take the limiting value of an equation as $\Delta x \to 0$. Using the lingo and terminology of Calculus, this is the derivative. Thus, for the parabola, $y = x^2$, Eqn. 1-16 and Eqn. 1-20 reduce to,

$$\boxed{\frac{\mathrm{d}\left[x^2\right]}{\mathrm{d}x}=2x} \qquad \text{Eqn. 1-21}$$

Let's play the same game with the Cubic Curve, $y = x^3$.

$$y + \Delta y = (x + \Delta x)^3 = x^3 + 3x^2\Delta x + 3x\Delta x^2 + \Delta x^3$$

Subtracting $\Delta y = x^3$ leaves,

$$\Delta y = 3x^2\Delta x + 3x\Delta x^2 + \Delta x^3$$

Then, dividing by Δx, produces the finite difference form of the derivative for $y = x^3$.

$$\frac{\Delta y}{\Delta x} = 3x^2 + 3x\Delta x + \Delta x^2$$

Taking the limit as $\Delta x \to 0$ produces the slope or derivative of the cubic curve,

$$\boxed{\frac{\mathrm{d}\left[x^3\right]}{\mathrm{d}x} = 3x^2} \qquad \text{Eqn. 1-22}$$

Are you ready to generalize (by induction)? When $y = x$, we know that the slope, m of the line equals 1 (or $1x^0$). When $y = x^2$ the slope is $2x^1$. When $y = x^3$ the slope is $3x^2$. By induction or by the binomial series (described below), when $y = x^n$, we find that its slope or derivative is given by (are you ready?),

$$\boxed{\frac{\mathrm{d}\left[x^n\right]}{\mathrm{d}x} = nx^{n-1}} \qquad \text{Eqn. 1-23}$$

This is one of the great results of Calculus!

1.5 The Binomial Theorem, Series and Distribution: Probability

Proving Eqn. 1-23 hinges on expanding $(x + \Delta x)^n$ by multiplying (i. e., un-factoring). Finding a general formula for $(x + \Delta x)^n$ is not too tough if n is an integer. Isaac Newton guessed that the same general formula works when n is not an integer even though it leads to an infinite series. The result (proven much later) is the **Binomial Theorem**,

$$\boxed{(p+q)^n = p^n q^0 + \frac{n}{1!}p^{n-1}q^1 + \frac{n(n-1)}{2!}p^{n-2}q^2 + \ldots + \frac{n!}{n!}p^0 q^n} \qquad \text{Eqn. 1-24}$$

Notice that I have used $(p + q)^n$ in place of $(x + \Delta x)^n$. I have not done this to confuse you although it might. There is a high probability that I have a secret motive (which I will reveal after I discuss a few beautiful features of the series).

The coefficients of the p's and q's in Eqn. 1-24 are given by the famous Pascal's Triangle in Figure 1-14. The number in the 1st line of the triangle is the coefficient of $(p + q)^0$; the numbers in the 2nd line are the

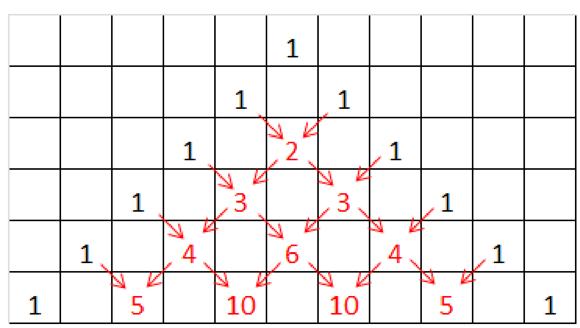


Figure 1-14 Pascal's Triangle.

coefficients of $(p + q)^1$, and the numbers in the n^{th} line are the coefficients of $(p + q)^{n-1}$ in Eqn. 1-24. Notice that these numbers are also symmetrical from left to right.

Pascal's triangle is constructed so that the external numbers consist of black 1's. Each interior red number is the sum of the two numbers diagonally above it.

The numbers in Pascal's Triangle are also the coefficients of the binomial series and also the number of combinations (teams) of *k* items selected from a pool of *n* items (recall Section 0.9 and Eqn. 0-38).

$$\boxed{\frac{n!}{k!(n-k)!} \equiv \binom{n}{k}} \qquad \text{Eqn. 0-38}$$

With the standard shorthand $\binom{n}{k}$, the binomial series, Eqn. 1-24, can be written as,

$$\boxed{(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}} \qquad \text{Eqn. 1-25}$$

Gambling and Probability: The Binomial Distribution and Coin Tossing

Gambling represents one of the great applications of the binomial distribution and should therefore represent one of the great joys or woes of life. The binomial distribution gives the probability of combinations of simple events with only two alternatives or choices, such as tossing a coin (heads or tails – disregarding the chance a coin will stand on its edge), the sex of a child (male or female – disregarding…), losing or winning the lottery, will it rain or not, will there be a flood or not. Phrased poetically,

To be or not to be, that is the binomial question!

Each term in Eqn. 0-38 gives the number of combinations of getting *k* (e. g., 8) heads out of *n* (e. g., 10) coin flips or, in more general terms, the number of combinations of *k* successes out of *n* tries no matter how likely or unlikely each combination is. Each term in the sum of Eqn. 1-25 gives the probability of *k* successes out of *n* tries and includes the possibility that heads may be more likely than tails.

Can you tell what letter we use to represent ***p***robability of a success for a single event? You guessed it, ***p***! And that is why I used *p* in Eqn. 1-25. But what then is *q*? Think hard! This is another SAT question. Either an event happens or it doesn't. The total probability is 100% or 1. So, $p + q = 1$ and

$$\boxed{p+q=1 \quad \Leftrightarrow \quad q=1-p} \qquad \text{Eqn. 1-26}$$

When Eqn. 1-26 is valid the left hand sides of Eqn. 1-24 and Eqn. 1-25 reduce to 1.

To verify that the binomial distribution governs probabilities of combinations of events with only two choices, consider how the sexes are distributed in a family with 3 children born at separate times (so we don't count identical twins). Table 1-1 shows that 3 birth events have $2^3 = 8$ possible arrangements. Only 1 arrangement each produces a family with 3 boys or a family with 3 girls but 3 arrangements each produce a family with 2 girls and 1 boy or a family with 1 girl and 2 boys. This corresponds to the 1 – 3 – 3 – 1 line of Pascal's Triangle. (Do you remember seeing this before in Table 0-5 [possible arrangements of flipping 3 coins]? It was worth repeating.)

Child\Combo #	1	2	3	4	5	6	7	8
1	B	B	B	B	G	G	G	G
2	B	B	G	G	B	B	G	G
3	B	G	B	G	B	G	B	G

Table 1-1 All possible arrangements of the sexes in a family of 3 children.

Now let's talk probability. This is where we can make or lose a lot of money. The probability of a single event (one coin toss) may or may not be 50%. If a coin is fair there is exactly a 50% chance of heads or tails on any one toss. Of course, a coin will be biased if one side is glued or slightly magnetized. Then the probabilities will no longer be 50-50. The probability that a newborn is a boy is slightly greater than 50%, but the probability of encountering women in nursing homes is much larger than 50% because on average, women live much longer than men. Finally, we all know that the probability of winning the lottery or flooding the New York City Subways (oops!) is much less than 50%, at least for any one person or on any one day.

Probability questions get complicated real fast. When $n = 3$ we just saw that there are $2^3 = 8$ arrangements. The number of arrangements increases to $2^{10} = 1024$ when $n = 10$, to $2^{20} = 1{,}048{,}576$ when $n = 20$, and to $2^{100} \approx 1.27(10)^{30}$ when $n = 100$. Figure 1-15 compares the binomial distributions for $n = 3$, 10, 20, and 100. You can see that the distribution become narrower and more sharply peaked so that extreme outcomes (far from the midpoint or mean) become less and less likely as n increases. Most people are surprised by how narrow the probability curve is when n is large.

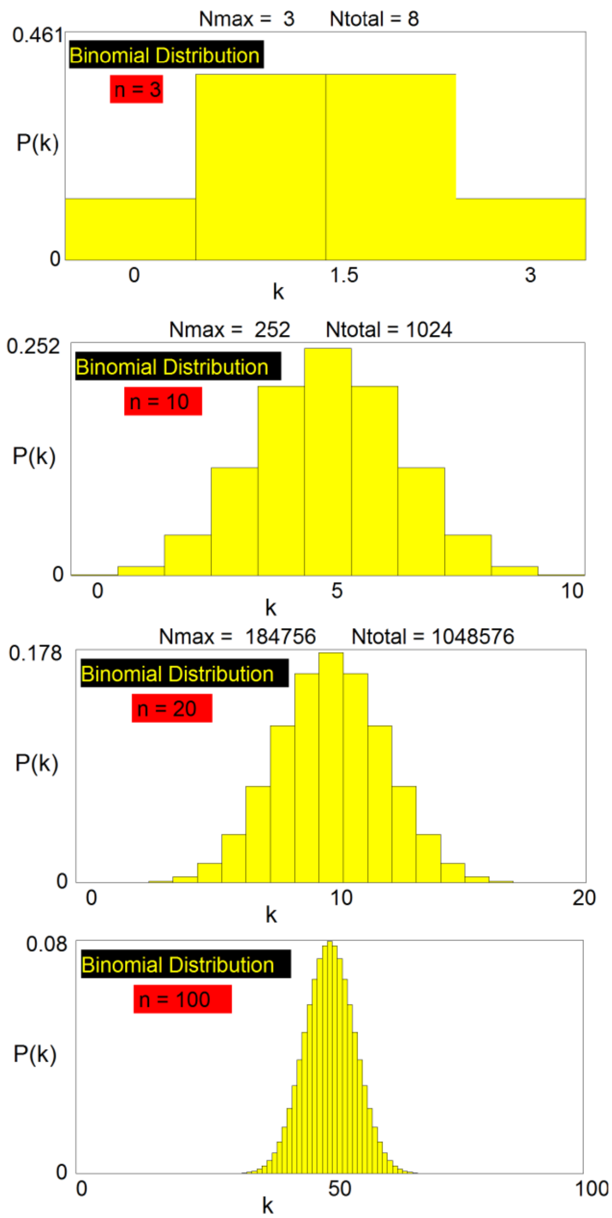


Figure 1-15 Binomial distributions for $n = 3$, 10, 20, and 100. The probabilities of extreme events (e. g., $k < 30$ or $k > 70$) for $n = 100$ are too small to see.

Whenever I teach Probability and

Statistics, I capitalize on the surprising narrowness of the binomial distribution for large n. I offer students the following bet. Flip a coin n = 100 times. If anywhere from 76 to 100 heads comes up I pay each student $1000. If 75 or fewer heads come up each student must pay me only $1.

That is 1000:1 odds! Even though some of the students suspect I am tricking them many are strongly tempted to take my bet. What they don't know is that I calculated their probability of winning using the binomial distribution, and found that is about 2 in 29 million! Of course, I never take any money so the entire exercise is called a thought experiment because otherwise I would be accused of a whole range of crimes and be fired. But I would be rich until my lawyer defended me!

Calculating the probability of getting more than 75 heads is a tough job without a program such as Excel. So, at the end of Chapter 6 I will show you how to calculate the approximate probability of such enormous numbers very quickly using results from Calculus. Right now, though let's consider two simple problems.

Problem: Calculate the number of arrangements and the probability of getting k = 8 or more heads out of n = 10 tosses of a fair (50-50) coin.
Solution: There are $2^n = 2^{10}$ = 1024 arrangements. Eqn. 0-38 shows there is

A: 1 arrangement of k = 10 heads,
B: 10 arrangements of k = 9 heads
C: 10×9/2 = 45 arrangements of k = 8 heads.
D: In sum, there are 64 arrangements of 8 or more heads out of a total of 1024. Assuming the coin is not biased the probability is 64/1024 or about 6.4%.

Problem: Calculate the number of arrangements and the probability of getting k = 16 or more heads out of n = 20 tosses of a fair coin.
Warning: Although 16 or more out of 20 may sound like 8 out of 10 the problem is different and the solution is different!
Solution: With 20 tosses there are 2^{20} = 1048576 arrangements (a bit over 1 million). There is 1 arrangement of 20 heads, 20 arrangements, of 19 heads, 20×19/2 = 190 arrangements of 18 heads, 20×19×18/(2×3) = 1140 arrangements of 17 heads, and, 20×19×18×17/(2×3×4) = 4845 arrangements of 16 heads for a total of 6206 arrangements out of 1048576. Thus, the probability of getting 16 or more heads out of 20 flips (or a minimum of 80% accuracy) is only about 6206/1048576 or 0.6%. Already the odds are less than 1:100.

When you flip a coin 100 times, the number of arrangements is $2^{100} \approx 10^{30}$, which is roughly the number of sand grains that would fit into the Moon. So, perhaps you see why the probability of getting more than 75 heads out of 100 flips is so low.

Problem: I sometimes imagined what it would be like to have a birthday on February 29. That only happens 1 day out of 1461 = 365×4 + 1. So, what is the probability that at least 2 out of 3 children in a family are born on February 29?
Solution: Remember that with 3 events there are 2^3 = 8 arrangements. Now p = 1/1461, which is quite small. So $p^3 << p^2$ and $(1 - p) \approx 1$, so we have

$$P(2)+P(3)=p^3+8p^2(1-p)\approx 8p^2=8/1461^2\approx 3.75/1{,}000{,}000$$

This is roughly 1 out of 300 thousand. Not impossible but pretty rare.

If you have learned anything from this short diversionary section, it is to not gamble because long ago, casino statisticians figured out all the odds and stacked them all against you. If you get a pleasure from the illusion of winning, simply send me what you would have bet and lost in a casino. You must admit it will go to a far more worthy cause. After all, the casinos are only after your money while I am trying in this book to do you some good. I will return to the subject of probability at the very end of the book, because it is essential to learn how to deal best with uncertainty in life.

1.6 Integrals and Derivatives are Inverses: The Fundamental Theorem of Calculus Illustrated by Acceleration, Velocity, Distance – Newton's Equations of Motion

This section begins by introducing force = *F*, velocity = *v*, acceleration = *a*, and displacement = *s*. Note: These are vector quantities (see Section 6.13) because they include direction as well as magnitude. For example, velocity is speed in a specified direction. Thus, the vertical velocity of a falling object is negative. By using the proper signs and components, we include the vector nature of these quantities implicitly until Section 6.13.

We then use Newton's Second Law of Motion to show that taking integrals and derivatives are inverse processes, much like multiplying and dividing are. Newton had a profound motive for inventing integrals and derivatives. He needed them to explain the motions of objects, including heavenly bodies, in terms of a fundamental principle. We know the famous simplification of his law that an object subjected to a force, *F*, accelerates in the direction of the force so that the force equals the mass, *m* (quantity of matter) of the object times the acceleration, *a*, or,

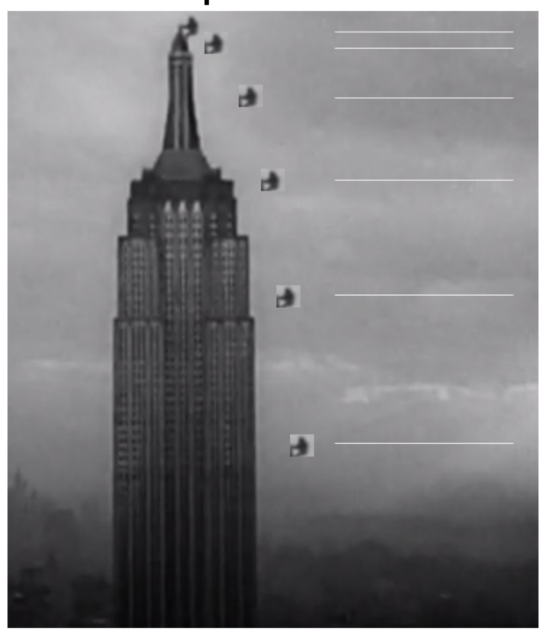


Figure 1-16 King Kong accelerates as he falls off the Empire State Building.

$$\boxed{F = ma} \qquad \text{Eqn. 1-27}$$

Acceleration (Eqn. 1-28) is the change of velocity, d*v* divided by the time interval, d*t*. That makes it both a fraction and a derivative. Velocity (Eqn. 1-29) is also a fraction and a derivative because it is defined as the change of distance, d*s* divided by d*t*.

$$\boxed{a = \frac{\mathrm{d}v}{\mathrm{d}t}} \qquad \text{Eqn. 1-28}$$

$$\boxed{v = \frac{\mathrm{d}s}{\mathrm{d}t}} \qquad \text{Eqn. 1-29}$$

Of course, you knew about acceleration, velocity, and distance long before you got to Calculus. Acceleration is the most difficult concept of the three, but you know what it is qualitatively. When King Kong fell off the Empire State Building he fell faster and faster with time. The white lines in Figure 1-16 show that King Kong fell faster and further each second (because of gravity).

A (hopefully) safer way to feel acceleration is in your car. Depress the gas pedal and speed increases, i. e., the car accelerates. To calculate the acceleration all you have to do is divide the change of the car's speed by the time it took to reach the new speed.

Our feeling for velocity is more quantitative. On the highway, the car travels with a velocity or speed, say of 62.5 mph (about 100 km per hr). We also have a quantitative idea of distance. Remember long car trips as a kid? "Are we there yet?" "No, kid, it will take us 4 hours." If you travel for 4 hours at 62.5 mph you will travel a distance of 4×62.5 = 250 miles (or 4×100 = 400 km). So, distance is equal to velocity times the time interval. But that also means **distance is the integral (with respect to time) of velocity.** Similarly, **velocity is the integral (with respect to time) of acceleration**.

$$\boxed{v = \int_0^t a\mathrm{d}t} \qquad \text{Eqn. 1-30}$$

$$\boxed{s = \int_0^t v\mathrm{d}t} \qquad \text{Eqn. 1-31}$$

Of course, you could ask the problem the other way around, namely if you travel 250 miles in four hours, what is your average speed? Then you would solve by dividing distance by the time interval or take the derivative of distance using Eqn. 1-29. To repeat, velocity is the derivative of distance and distance is the integral of velocity. Multiply by time or integrate over time and you get distance from velocity. Divide by the time interval or differentiate with respect to time and you get velocity from distance. Get the point? The point is,

Taking derivatives (differentiating) and taking integrals (integrating) are inverse processes just as are dividing and multiplying.

When Newton found that speed is the derivative of distance and distance is the integral of speed he realized something fundamental that no one else playing with areas and slopes had quite put together. His brainstorm was The Fundamental Theorem of Calculus - **The integral of a function is its antiderivative! Thus, the integral is the opposite of the Derivative**!

$$\boxed{F(x) = \int_0^x f(x)\mathrm{d}x \quad \Leftrightarrow \quad \frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x)} \qquad \text{Eqn. 1-32}$$

Figure 1-17 illustrates these concepts. The left panel depicts acceleration. Its integral over time (the green area) is the velocity. In this case $v = t$ is linear. The integral of velocity over time (the green area in the center panel) is distance. In this case, $s = ½t^2$ is a parabola. Conversely, the slope, or derivative of the velocity with respect to time is the acceleration, and the derivative of the distance with respect to time is the velocity.

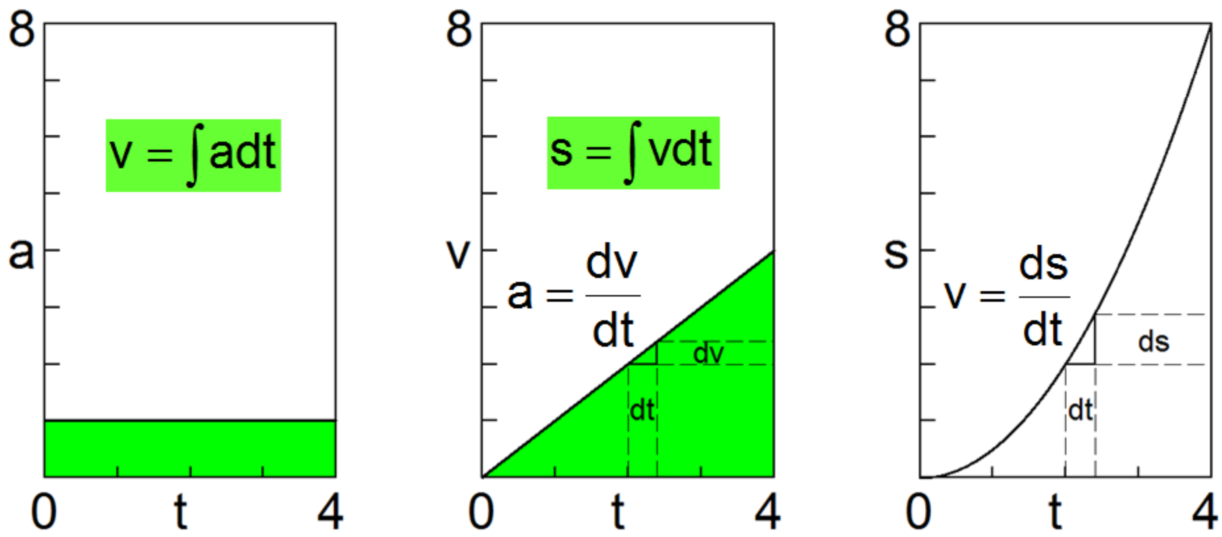


Figure 1-17 Acceleration, *a*, velocity, *v*, and distance, *s* vs. time. Green areas represent integrals; slopes are the derivatives. The integral of distance vs. *t* has no simple meaning.

1.7 A Giant Step Backwards: Systems and Numerical Solutions

Some functions are so difficult that even I cannot find a general expression for a derivative or integral. If this happens, do you think I panic? You bet I do! But once I regain my composure, I remember that I can certainly calculate their values quite accurately.

Rule: Whenever you get lost trying to determine derivatives and integrals always feel free to approximate.

Numerical techniques are essential for evaluating integrals and derivatives of difficult functions, and are used all the time on the computer. They often amount to taking a giant step backwards by reversing the approach of the Calculus and using a finite value for *Δx*.

Therefore, before *we* proceed, try to solve two simple problems just to show yourselves that you can *evaluate* the derivative and the integral of any function, even if you have no idea whatsoever of how to find their general expressions.

Problem: Approximate the derivative of *sin(x)* when *x = 60°* (π/3 radians).
Reminder: The derivative is just like the ratio of differences,

$$\frac{dy(x)}{dx} \approx \frac{y(x+\Delta x)-y(x)}{\Delta x}$$

You must use radians. Then $x = \pi/3 \approx 1.0472$. You are free to choose any value of radians for *Δx*. Of course, you would be wise to choose a small value. I will choose *Δx* = .01.
Solution

$$\frac{\sin(x+\Delta x)-\sin(x)}{\Delta x} \approx \frac{\sin(1.0572)-\sin(1.0472)}{0.01} \approx \frac{0.870983-0.866027}{0.01} \approx 0.4957$$

This is pretty darned close, considering that the value of the derivative is 0.5.

Problem: Approximate the integral of $v = 1/(1 + t)$ for $0 \le t \le 4$.
Reminder: The integral is just like the sum of the areas of rectangles. To increase accuracy, use skinny rectangles so choose a small value of Δt. But the smaller Δt, the more rectangles or calculations we must make. This calls for a compromise. For starters, set $\Delta t = 0.5$, so that there are 8 rectangles, and pray the answer won't be so bad.
Solution: The first rectangle extends from $t = 0$ to $t = 0.5$. At $t = 0.5$, the height of the first rectangle, $v = 1/(1 + t) = 1/(1 + 0.5) = 0.667$. The area of the first rectangle is therefore, $A_1 = v\Delta t = 0.333$. The heights of all the rectangles are given in Table 1-2. The total area of the rectangles is 1.429. This result is not too bad because the actual value of the integral rounds off to 1.609.

t	v=1/(1+t)
0	1.000
0.5	0.667
1	0.500
1.5	0.400
2	0.333
2.5	0.286
3	0.250
3.5	0.222
4	0.200
ΣvΔt =	1.429

Table 1-2 Numerical solution of v vs. t.

Want a more accurate result? You'll have to do more work by taking a larger number of narrower rectangles. Of course this isn't difficult if the computer does the work (i. e., the case for slavery).

Whenever I can't find a general expression for a derivative or integral, but need some answer, I always take a giant step backwards, calculate a finite difference ratio or Riemann Sum on the computer or in Excel, and graph so I can see.

Predictions and Systems

Predicting is a key goal of science. Weather prediction has improved markedly over the past half century because of measurements from satellites and radar, and because of the computer and the scientists who write the models that are used to make numerical predictions using Newton's law of motion, etc.

When making predictions it is useful to return to the concept of a System, introduced in Section 0.8 and illustrated in Figure 1-18. Unless we accept Creationism, the rate of change of storage of any quantity (such as mass or money) in any system is equal to the rate of input, R_{in} from the world outside the system minus the rate of output, R_{out} to the world outside the system and might be called the...

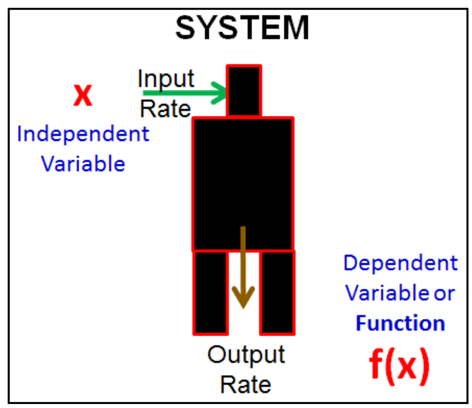


Figure 1-18 A system with input x and output f(x), a function of the system.

Fundamental Equation of Systems

(Rate of) Change of System Storage = Input (Rate) – Output (Rate)

$$\boxed{\frac{dy(t)}{dt} = R_{in} - R_{out}} \qquad \text{Eqn. 1-33}$$

In this section, t is the independent variable (in place of x) because predictions usually involve time. To predict future values of any quantity in a system, replace the derivative on the left hand side of Eqn. 1-33 with its finite difference form from Eqn. 1-16 and rearrange to solve for the future value, $y(t + \Delta t)$ at time, $t + \Delta t$. This defines the...

Fundamental Equation of Change

New Value = Old Value + Rate of Change × Interval

$$\boxed{y(t+\Delta t)\approx y(t)+\frac{\mathrm{d}y(t)}{\mathrm{d}t}\Delta t=y(t)+\left(R_{\text{in}}-R_{\text{out}}\right)\Delta t} \qquad \text{Eqn. 1-34}$$

In Eqn. 1-34, Δt is the time step, the independent variable that we are free to choose (but we keep Δt small for accuracy); $y(t)$ is the present value of y, $y(t+\Delta t)$ is the future value.

Now let's examine the finite difference prediction technique called <u>iteration</u> using the Fundamental Equation of Change

Forecasting by Iteration

1: Substitute the present value of $y(t)$ and its rate of change, dy/dt (or input rate minus output rate) in Eqn. 1-34 to find the future value, $y(t + \Delta t)$.
2: Update Eqn. 1-34 by replacing y and dy/dt at time t with the new values at time $t + \Delta t$.
3: Repeat (iterate) steps 1 and 2 as far into the future as needed.

Because iteration is cumulative, it is like integrating, as you will soon see. Smaller steps usually yield more accurate results, but then it takes more steps to reach the final time or point. This was a formidable problem before the computer but with growing computer power it is much, much less of a problem.

Let's use Eqn. 1-34 to forecast your savings in 10 years if you start with \$1.00 and earn 50% interest per year, assuming interest is paid at the end of each year. Results are shown in Table 1-3 and graphed in Figure 1-21, but we will manually take the first few steps now. In year 0 you start with $y(0)$ = \$1. At the end of year 0, after a 1 year time step, Δt = 1, you earn 50% on your dollar or \$0.50. So at the start of year 1 you have $y(1)$ = \$1.50.

year	Wealth	Interest
0	1.00	0.50
1	1.50	0.75
2	2.25	1.13
3	3.38	1.69
4	5.06	2.53
5	7.59	3.80
6	11.39	5.70
7	17.09	8.54
8	25.63	12.81
9	38.44	19.22
10	**57.67**	

Table 1-3 Wealth and interest vs. time.

$$y(\Delta t)\approx y(0)+\frac{\mathrm{d}y(0)}{\mathrm{d}t}\Delta t=1+0.5\times1=1.5$$

At the end of year 1 you earn 50% on \$1.50 = \$0.75, so $y(2)$ =

$$y(2)\approx y(1)+\frac{\mathrm{d}y(1)}{\mathrm{d}t}\Delta t=1.5+0.5\times1.5=2.25$$

At the end of year 2 you earn 50% on your \$2.250 or \$1.125, so $y(3)$ =

$$y(3)\approx y(2)+\frac{\mathrm{d}y(2)}{\mathrm{d}t}\Delta t=2.25+0.5\times2.25=3.375$$

After 10 years you will have \$57.67.

When savings or wealth is graphed as a function of time (as in Figure 1-19) you can see that it grows slowly at first but then grows faster and faster as time continues. It is an example of exponential growth. (An exponential function changes at a rate proportion to itself.) We will spend much time with exponential functions because so many processes and phenomena either grow or shrink exponentially.

We will return to treat systems and their changes in Chapter 5 because the Equation of Change is the key to solving difficult differential equations and forms the basis for computer forecasts.

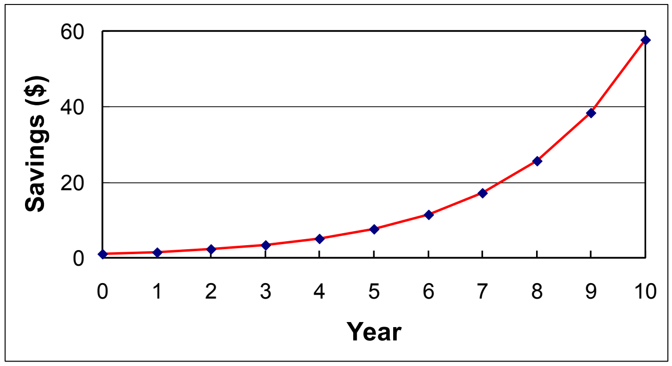


Figure 1-19 Savings or wealth vs. time.

1.8 What We Have Done and What We Haven't Done

In this chapter we have used some inductive reasoning but have calculated just about everything – the approximate value of integrals, derivatives, and differential or finite difference equations. We have calculated, calculated, calculated. We have acted as computers. There is great strength in that. After all, that is how weather is forecast. But it is tiring to have to calculate the approximate value of every single new integral and derivative.

So far, we have only touched Calculus but we have seen a hint of its power. Calculus is more abstract, but also more general. Using the Method of Exhaustion we found that the derivative of the parabola, $y = x^2$ is $2x$ and its integral is $\frac{1}{3}x^3$. Calculus will give us many derivatives and integrals of many more functions. Admittedly, many features of nature, such as weather patterns or human faces are too erratic and complex to be reduced to simple functions. But once Calculus gives the general formulas for the derivative and integral of a function, a young child can calculate them.

When Calculus cannot solve a problem exactly, it often gives techniques for calculating approximate solutions. Important examples are Taylor Series and Fourier Series (see Chapter 4), which enable us to 1: approximate any complex function and, 2: simplify the calculations we program into calculators and computers.

Calculus is not only a powerful mathematical tool, it is one of the great intellectual adventures of the human race – a triumph – a tour de force. It opened the door to many of the great scientific and technological discoveries and breakthroughs that have extended lifespans and added to the quality of life as well. Yes, we have not always used science and technology wisely or well. But we now live long enough to kill ourselves rather than tremble ignorantly and helplessly in the dark when famine, plague, storms, or wild beasts strike. And we have extended the success and impact of Calculus outside of science and mathematics to the Social Sciences. Calculus has been an integral part of our self-empowerment and self-realization.

In this Chapter you have embarked on the great adventure of Calculus. When you get through this book, which might not always feel like recreation, you will have re-created the genius of others. You will have become the genius you always secretly knew you were.

May your learning curve, like your interest and wealth, grow exponentially!

1.X (EXTRA) Mortgaging your Life and the Geometric Series

Originally, this page was empty. I decided to fill it because 1: I have mortgaged my life to pay for producing this book, 2: the equation for mortgage payments is based on Eqn. 1-2 for the sum of a geometric series and, 3: the financial crisis of 2008 could have been avoided if millions of people could have calculated what they could afford.

Let's get started! Consider fixed rate mortgages because the others are 1: harder to calculate, 2: too often rip-offs. The job – find the monthly cost, C, of a fixed rate mortgage with principal, P, duration, n months and monthly interest rate, x.

Perhaps the best way to do this is to calculate your debt, D_k after k months. The moment you sign, month, k = 0, your debt, $D_0 = P$. That's the easy one. By month k = 1, the debt has grown to $D_0 \cdot (1 + x) = P \cdot (1 + x)$ the instant before you make your first payment. When you make your payment an instant later, your debt is reduced by C to,

$$D_1 = D_0(1+x) - C = P(1+x)^1 - C$$

At month, 2, prior debt is increased by the same factor, $(1 + x)$ and then reduced by C.

$$D_2 = D_1(1+x) - C = P(1+x)^2 - C\left[(1+x)^1 + 1\right]$$

Watch as a pattern emerges. The debts at months, 3 and 4 are,

$$D_3 = D_2(1+x) - C = P(1+x)^3 - C\left[(1+x)^2 + (1+x)^1 + 1\right]$$

and

$$D_4 = D_3(1+x) - C = P(1+x)^4 - C\left[(1+x)^3 + (1+x)^2 + (1+x)^1 + 1\right]$$

The equation keeps getting longer. Is all hope lost? Absolutely not! The terms in brackets constitute a geometric series and Eqn. 1-2 gives its sum. Thus, in general,

$$\boxed{D_k = P(1+x)^k - C\sum_{i=0}^{k-1}(1+x)^i = P(1+x)^k - C\frac{(1+x)^k - 1}{x}} \qquad \text{Eqn. 1-35}$$

At the n^{th} payment debt the mortgage is done, so debt $D_n = 0$. Solving for C yields,

$$\boxed{C = \frac{xP(1+x)^n}{(1+x)^n - 1}} \qquad \text{Eqn. 1-36}$$

Problem: Find the monthly cost and total interest paid for a P = \$1,000,000 mortgage for 30 years at 5% annual interest.

Solution: There are $n = 12\times30 = 360$ monthly payments with monthly interest rate, $x = 0.05/12 \approx 0.00417$. Substituting in Eqn. 1-36 yields a monthly payment of,

$$C = \frac{xP(1+x)^n}{(1+x)^n - 1} = \frac{0.00417\times10^6\times(1.00417)^{360}}{(1.00417)^{360} - 1} \approx \$5388.62$$

Multiply by 360, subtract 1 million and total interest of the mortgage is \$932,557.80!

Warning: Mortgage payments should be no more than ¼ of monthly income, so if you want that million dollar mortgage you better earn $4\times5388.62\times12 \approx$ \$258,630 per year!

CHAPTER 2: DERIVATIVES AND TRANSCENDENTAL FUNCTIONS

If you've reached this point, knowing all that has come before, you're in great shape to get into serious Calculus. **Quiz**: Do you remember the derivative of x^n???

$$\boxed{\frac{\mathrm{d}[x^n]}{\mathrm{d}x} = nx^{n-1}} \qquad \text{Eqn. 1-23}$$

Your professor knows that you know this simple derivative, so don't expect to find it on your first Calculus test, unless you have to derive it! Much more likely, you will have to find the derivatives of much more difficult functions like sin*(x)*, log*(x)*, and e^x. This will demand that you finally learn your algebra, geometry and trigonometry. It will also test your limits. **So, let the games begin!**

Why do derivatives come before integrals? (Did you think of asking that question?) Even though approximating integrals with Riemann sums is straightforward and easy, it is much easier to find exact equations for derivatives than for integrals, just as multiplying is easier than dividing or squaring a number is easier than finding its square root. In fact, the only way to find the integral of many functions is to proceed in reverse by guessing an answer and taking its derivative until with luck, sweat, and toil it matches the function.

2.1 The Derivative – Know Your Limits I

I hope you remember that the derivative of a function, *y*(*x*) is defined by Eqn. 1-16 as,

$$\boxed{\frac{\mathrm{d}y(x)}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\Delta y(x)}{\Delta x}} \qquad \text{Eqn. 1-16}$$

A small but significant question is how to calculate $\Delta y(x)$. We have 3 simple choices.

$$\Delta y(x) = \begin{cases} y(x+\Delta x) - y(x) & \rightarrow \quad \text{Forward Difference} \\ y(x+\frac{1}{2}\Delta x) - y(x-\frac{1}{2}\Delta x) & \rightarrow \quad \text{Center Difference} \\ y(x) - y(x-\Delta x) & \rightarrow \quad \text{Backward Difference} \end{cases} \qquad \text{Eqn. 2-1}$$

Remember! Colors are used simply to highlight.

The center difference is usually the most accurate because it is symmetrical, but it has the disadvantage that it involves past, present, and future. To calculate the present rate of change of the function you must know both the function's past and future values.

The forward and backward differences are less accurate but each involves only two times. If you use the forward difference form of the derivative, you only need to know the present values to determine the future. This is the way to make a forecast. The backward difference tells us about the past, but who cares because we all have 20:20 hindsight.

So, it is decided! Unless some emergency arises, use forward differences!

Sines of the Times – Know Your Limits II

Let's get transcendental and see if we can find the derivative of *sin(x)*. First, however, you might want to know what is so important about the sine and its derivative? Remember that the sine or cosine curves are waves, and waves are omnipresent in nature. Not only are there waves in the ocean, tsunamis included, but we hear sound waves, and transmit and receive electromagnetic waves over our cell phones.

The sine wave is not the only wave form. Breaking ocean waves and river meanders are examples of waves with large folds or asymmetries that do not match simple sine waves (see Figure 2-1). However, the sine wave is both mathematically and physically the simplest wave and that is one crucial reason we spend so much time with it. Another crucial reason is that we can build most complex waves from simple sines and cosines using a technique called Fourier Analysis that you will learn in Chapter 4.

Figure 2-1 Breaking waves (left) and river meanders (center). Their shapes, shown in the cartoon (right), don't match sine waves.

In order to understand the dynamics of waves we absolutely must know the derivative and integral of the sine and cosine. Let's find the derivative of sin(*x*). **Warning**: It is not easy. The first time I tried to find it, I made a good start. I took the difference between sin(*x*+Δ*x*) and sin(*x*) and divided by Δ*x*, or

$$\frac{\mathrm{d}[\sin(x)]}{\mathrm{d}x}=\lim_{\Delta x\to 0}\frac{\sin(x+\Delta x)-\sin(x)}{\Delta x}$$

I then substituted in the multiple angle formula for sines (Eqn. 0-12),

$$\sin(x+\Delta x)=\sin(x)\cos(\Delta x)+\sin(\Delta x)\cos(x)$$

This left,

$$\frac{\mathrm{d}[\sin(x)]}{\mathrm{d}x}=\sin(x)\cdot\lim_{\Delta x\to 0}\left[\frac{\cos(\Delta x)-1}{\Delta x}\right]-\cos(x)\cdot\lim_{\Delta x\to 0}\left[\frac{\sin(\Delta x)}{\Delta x}\right]$$

This is where I got stuck because I didn't know my limits. After a good cry, I realized that I had learned the hard way a lesson that generations of Calculus students have had to learn. In order to get anywhere in Calculus you must know your limits!

First, I'll **tell** you the limits of the terms in the square brackets so that we can find the derivative of sin(x). Then, we'll **derive** those limits.

The limit for the sine is,

$$\boxed{\lim_{\Delta x \to 0}\left[\frac{\sin(\Delta x)}{\Delta x}\right] = 1} \qquad \text{Eqn. 2-2}$$

The limit for the cosine is,

$$\boxed{\lim_{\Delta x \to 0}\left[\frac{\cos(\Delta x) - 1}{\Delta x}\right] = 0} \qquad \text{Eqn. 2-3}$$

Using these limits, we find that the derivative of the sine is the cosine!

$$\boxed{\frac{\mathrm{d}[\sin(x)]}{\mathrm{d}x} = \cos(x)} \quad \text{or} \quad \boxed{\frac{\mathrm{d}[\sin(\theta)]}{\mathrm{d}x} = \cos(\theta)} \qquad \text{Eqn. 2-4}$$

This marvelous result should not be a complete mystery if you overlay the graphs of sin(θ) and cos(θ), as in Figure 2-2.

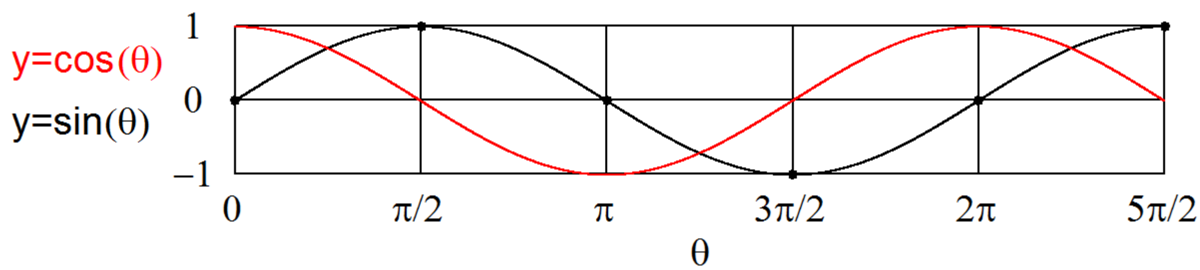


Figure 2-2 Sine and Cosine waves. When the sine curve is rising the cosine curve is positive.

In Figure 2-2, compare the *slope* of the sine curve with the *height* of the cosine curve and note the following properties:

1: At θ = 0 the sine curve starts off heading up to the right with its steepest slope (largest positive derivative). At $\theta = 0$ the cosine has its maximum value [cos(0) = 1].

2: As θ increases, the *slope* of the sine curve becomes gentler (the *derivative* decreases), until you reach the top where θ = $\pi/2$, and the sine curve is level (derivative is zero). Similarly, as θ increases the cosine decreases until θ = $\pi/2$ where cos(π/2) = 0.

3: Beyond θ = $\pi/2$ the sine curve begins heading down so that its derivative is negative. Similarly, beyond θ = $\pi/2$ the cosine is negative.

Now, using Figure 2-2 and the same reasoning, see if you can guess that d[cos(*x*)]/d*x* =

$$\boxed{\frac{\mathrm{d}[\cos(x)]}{\mathrm{d}x} = -\sin(x)} \qquad \text{Eqn. 2-5}$$

If you guessed correctly then there is great hope for you! If you included the minus sign, you are a potential genius! But don't get too saucy yet! We've got a long way to go!

Proving that the Limit of sin(*x*)/*x* = 1 as *x* → 0

Well, we've beaten around the bush long enough. Now, let's find our limits. First, prove Eqn. 2-2, that sin(*x*)/*x* → 1 as *x* → 0. Figure 2-3 includes a circle of radius, *r* = 1. Then arc length, *x* = angle, *x*. For any angle *x*, the shortest distance to the horizontal line is the line equal to sin(*x*) because it is perpendicular to the horizontal line. The line for tan(*x*) is longest because it is most oblique. Arc length, *x* is intermediate. The inequality for this relation is,

$$\sin(x) < x < \tan(x)$$

Next, invert each term to reverse the direction of the inequality and multiply each term by sin(*x*).

$$1 > \frac{\sin(x)}{x} > \cos(x)$$

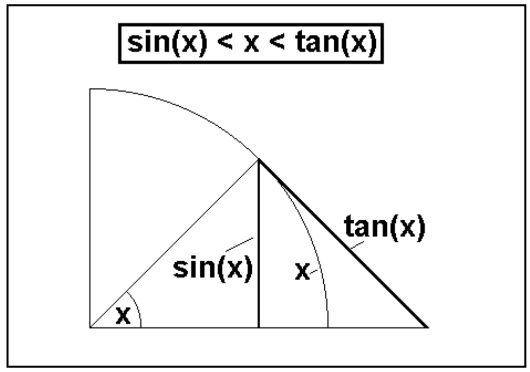


Figure 2-3 Sin(x) < x < tan(x)

Because cos(*x*) → 1 as *x* → 0 sin(*x*)/*x* is squeezed between 1 and 1 as *x* → 0. That doesn't leave much breathing room, does it? So, it proves Eqn. 2-2, namely that as x → 0 the limit of sin(*x*)/*x* = 1.

We now use trigonometry and the Binomial Series to prove Eqn. 2-3, that [1-cos(*x*)]/*x* → 0 as *x* → 0. First use Eqn. 1-10 to replace $\cos^2(x)$ with $[1 - \sin^2(x)]$ in Eqn. 2-6.

$$\frac{1-\cos(x)}{x} = \frac{1-\sqrt{1-\sin^2(x)}}{x} \qquad \text{Eqn. 2-6}$$

Radicals (square root terms) are always tough to play with, so it is best to get rid of them. There are at least two standard procedures, 1: Square or, 2: Use the Binomial Series expansion and throw out all tiny terms. A typical technique is to keep the two largest terms. When x is small, successive terms in the series get successively smaller because they have higher orders of *x* (x^n, for *n* > 1)

Remember also that sin(*x*) ≈ *x* when *x* is small. Thus, when *x* << 1 (i. e., when *x* is tiny), we can approximate $[1-\sin^2(x)]^{0.5}$ with the first two terms of the Binomial Series expansion.

$$\sqrt{1-\sin^2(x)} \approx \sqrt{1-x^2} = 1 - \frac{1}{2}x^2 + \frac{1}{2}\left(-\frac{1}{2}\right)\frac{\left(x^2\right)^2}{2!} - \ldots \approx 1 - \frac{1}{2}x^2$$

Substituting into Eqn. 2-6 yields,

$$\lim_{x\to 0}\left(\frac{1-\cos(x)}{x}\right) \approx \lim_{x\to 0}\left(\frac{1-\left(1-\frac{1}{2}x^2\right)}{x}\right) \approx \lim_{x\to 0}\left(\frac{1}{2}x\right) = 0$$

This proves Eqn. 2-3. Having proven the limits for Eqn. 2-2 and Eqn. 2-3 we are confident that d[sin(x)]/dx = cos(x) and d[cos(x)]/dx = -sin(x).

2.2 Derivatives of Derivatives

The discoveries of tiny worlds seen through microscopes and of infinitesimal numbers through series and Calculus excited the great satirist, Jonathan Swift to write a little poem,

So nat'ralists observe, a flea
Hath smaller fleas that on him prey;
And these have smaller fleas to bite 'em.
And so proceeds Ad infinitum."

The theme of this section is that, somewhat like fleas, derivatives have their own derivatives, and so on Ad infinitum! Higher order derivatives do not merely represent mathematical games, they has real and important applications. I'll give you two examples.

Example 1: The most popular hike up Mount Washington in New Hampshire starts from Pinkham Notch with a leisurely walk along a slightly sloped path. The slope or derivative, dy/dx, is small. Very soon though, the path steepens, In other words, dy/dx increases. After an hour or two you reach the base of Tuckerman's Ravine. Then the path up the walls of the Ravine gets scary steep very fast (dy/dx gets very large very fast). Once atop the ravine the slope for the path to the summit becomes gentler, so dy/dx decreases.

The rate of change of the slope described in this example is the derivative of the derivative of height. The derivative of the derivative is called the **Second Derivative**. When the derivative increases (decreases) the second derivative is positive (negative).

Example 2: Acceleration is a second derivative because it is the derivative of velocity, v with respect to time, t, (Eqn. 1-28) which is the derivative of distance, s with respect to time (Eqn. 1-29). When velocity increases, acceleration is positive.

$$\boxed{a \equiv \frac{\mathrm{d}[v]}{\mathrm{d}t} \equiv \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\mathrm{d}s}{\mathrm{d}t}\right]} \qquad \text{Eqn. 2-7}$$

Acceleration may be a little flea, but it is as important as it feels. When we travel we want the car, plane, or train to accelerate. People are hired and unemployment falls when the economy accelerates (i. e., picks up); people are fired and unemployment rises when the economy decelerates (i. e., falls into recession).

Acceleration relates to something even more important than the economy – bungee jumping – because it will make you think about sine and cosine waves! You begin on the platform. Jump and immediately your heart lodges in your mouth as you accelerate downward! But soon, though you are still falling, you feel like someone is pulling you up, and indeed, though you are moving down you are accelerating up and after bottoming, you are heading up again like a sling shot. As the bungee cord slackens gravity takes over and starts accelerating you downward. The oscillations rapidly damp out, so that after a few bounces you have spent loads of money the ride is over, and you get dragged up like a fish. But if the bungee cord were perfectly elastic and there were no air resistance, you would go on bouncing forever and ever, just like a cosine wave.

Think of the everlasting bungee jump mathematically. Your height (y) is a cosine wave where the x axis represents time (recall Figures 0-21 and 2-2). Near the top of the wave, the cosine or height is positive while the acceleration is negative (downward) and vice versa. Thus the second derivative or acceleration is the negative of the function (height). Surfers knew (or felt) all this decades ago.

An interesting thought may now occur to you if you compare the derivatives of sin(*x*) and cos(*x*). Since the derivative of sin(*x*) is cos(*x*) is and the derivative of cos(*x*) is -sin(*x*), then the second derivative of sin(*x*) is -sin(*x*)!

$$\frac{\mathrm{d}[\sin(x)]}{\mathrm{d}x} = \cos(x) \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\mathrm{d}[\sin(x)]}{\mathrm{d}x}\right] = \frac{\mathrm{d}[\cos(x)]}{\mathrm{d}x} = -\sin(x)$$

Thus the second derivative of the sine is a repeat of the original function only with a minus sign. Something cyclical is going on, which should not surprise you because the sine and cosine curves are waves, and waves are cyclical and repetitive by nature.

The *second derivative* is always written in the following way,

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\mathrm{d}y}{\mathrm{d}x}\right] \equiv \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}} \qquad \text{Eqn. 2-8}$$

Thus, the second derivative of sin(*x*) is,

$$\frac{\mathrm{d}^2[\sin(\mathrm{x})]}{\mathrm{d}x^2} = -\sin(x)$$

Some of you might wonder if there is a mystical significance about why the 2 appears **before** the *y* in the numerator and **after** the *x* in the denominator of the second derivative. There is none at all! The rationale is clear when we do the finite difference form of the second derivative. Remember that we write the forward difference of *y*(*x*) as,

$$\Delta y(x) = y(x+\Delta x) - y(x)$$

By shifting *x* forward a tiny step, +Δx, we write the forward difference of *y*(*x*+Δ*x*) as

$$\Delta y(x+\Delta x) = y(x+2\Delta x) - y(x+\Delta x)$$

With these steps, we are ready to write the second derivative in finite difference form. Remember that Δ*x* is a constant although we can give it any value we want. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\mathrm{d}y}{\mathrm{d}x}\right] \approx \frac{1}{\Delta x}\Delta\left[\frac{y(x+\Delta x)-y(x)}{\Delta x}\right] = \frac{1}{(\Delta x)^2}\left[\Delta y(x+\Delta x) - \Delta y(x)\right]$$

Using the values of Δ*y*(*x*) and Δ*y*(*x* + Δ*x*) given just above the second derivative becomes,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\mathrm{d}y}{\mathrm{d}x}\right] \approx \left[\frac{[y(x+2\Delta x)-y(x+\Delta x)]-[y(x+\Delta x)-y(x)]}{(\Delta x)^2}\right]$$

Grouping terms on the right leads to the forward difference form of the 2nd derivative,

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\mathrm{d}y}{\mathrm{d}x}\right]=\frac{\mathrm{d}^2y}{\mathrm{d}x^2}\approx\frac{y(x+2\Delta x)+y(x)-2y(x+\Delta x)}{(\Delta x)^2}} \qquad \text{Eqn. 2-9a}$$

If we back up by Δx, we get the center difference form of the 2nd derivative,

$$\boxed{\frac{\mathrm{d}^2y}{\mathrm{d}x^2}\approx\frac{y(x+\Delta x)+y(x-\Delta x)-2y(x)}{\Delta x^2}} \qquad \text{Eqn. 2-9b}$$

These long expressions are more important than you might think. They are used to calculate second derivatives such as acceleration in computer programs. They also justify the placement of the 2's in Eqn. 2-8 and Eqn. 2-9 because the numerator, d^2y is a difference of a difference while the denominator, $\mathrm{d}x^2$ is the square of Δx (as $\Delta \mathrm{x} \to 0$).

Problem: Approximate the 2nd derivative of the parabola, $y = x^2$.
Solution: Use Eqn. 2-9.

$$\frac{d^2\left[x^2\right]}{dx^2}\approx\frac{(x+2\Delta x)^2+(x)^2-2(x+\Delta x)^2}{(\Delta x)^2}=\frac{\left[x^2+4x\Delta x+4\Delta x^2\right]+x^2-\left[2x^2+4x\Delta x+2\Delta x^2\right]}{(\Delta x)^2}$$

When you get rid of all the terms that cancel, this long expression reduces to 2!

"What? It can't be that simple!"

"Oh, yes it can!"

"But is it correct?"

"Sure! Calculating the 2nd derivative of $y = x^2$ takes only two tiny steps for mankind."

$$\frac{\mathrm{d}x^2}{\mathrm{d}x}=2x \quad\Rightarrow\quad \frac{\mathrm{d}^2x^2}{\mathrm{d}x^2}=\frac{\mathrm{d}(2x)}{\mathrm{d}x}=2$$

Voila!

The meaning of the second derivative, $\mathrm{d}^2y/\mathrm{d}x^2 = 2$ of the parabola, $y = x^2$ is that its slope, s, increases at a constant rate of 2. When $x = 0$, $s = 0$; when $x = 1$, $s = 2$; when $x = 2$, $s = 4$; and when $x = 3$, $s = 6$. So s increases by 2 when x increases by 1.

Incidentally, a parabola is the trajectory that your junky old car takes when you push it off a cliff. Just as the parabola's slope increases at a constant rate, gravity accelerates the car downward at a constant rate. In one second it is falling at 10 $\mathrm{m{\cdot}s^{-1}}$ or about 20 mph. After two seconds it is falling at 20 $\mathrm{m{\cdot}s^{-1}}$, or about 40 mph. After 3 seconds it is falling at 30 $\mathrm{m{\cdot}s^{-1}}$, or about 60 mph. So, in its swan song appearance, your old car can go from 0 to 60 mph in 3 seconds. If you chart out its path, lo and behold, you will see that it is a parabola!

2.3 A New Interest – Exponentials

Warning: In this section I am switching to variable, *t* for time in place of x (since so many problems involving exponentials involve time), but I will switch to x whenever I want!

The sine and cosine are functions that are equal to minus their *second* derivatives. This might get you to ask the question, "Is there any function that equals its own derivative?" Of course there is! The **exponential function, e^t** is the function that equals its own derivative.

A slightly more general question is, "Is there any function that is proportional to its own derivative?" Once again, the answer is, "of course there is!" This condition is expressed mathematically by the differential equation,

$$\boxed{\frac{dy}{dt} = ky} \qquad \text{Eqn. 2-10}$$

The **exponential function, e^{kt}** is the function that is proportional to its own derivative, and is the solution to Eqn. 2-10.

$$\boxed{y(t) = y(0)e^{kt}} \qquad \text{Eqn. 2-11}$$

This may not help you very much because at this point you don't know how to solve Eqn. 2-10 and you may not know what ***e*** is. So I will give you a starting hint. ***e*** is a constant number with the approximate value

$$e \approx 2.71828\ldots \qquad \text{Eqn. 2-12}$$

So what? Wouldn't it be better to know what this number means and where it comes from, and why we consider it to be so very important?

The exponential function should generate a lot of interest. Why? Imagine that you earn 100% interest on your savings at the end of each year. (That is galaxies more than banks pay, but about what they charge for credit card loans.) If you started with \$10, you would earn \$10 interest at the end of the year. You would start the second year with \$20 and earn another \$20 at the end of that year, and so on. Thus, savings at 100% interest is a function that is equal to its own derivative.

If the annual interest rate is not 100%, the rate of growth is merely proportional to the principal. Would you like to know other situations where the function is proportional to its own derivative or should I leave you alone (a loan)? Many natural phenomena change at rates proportional to themselves. One example is unopposed reproduction, such as the early stages of population growth of cockroaches that invade a building. Another is the chain reaction in a nuclear explosion. Exponential functions grow exponentially.

Some exponential functions decay exponentially - the decay of radioactive elements not smashed together in a bomb, the loss of power in the transmission of light through long optical fibers or of electrical currents through long wires, or the loss of your wealth when the inflation rate is greater than the interest rate.

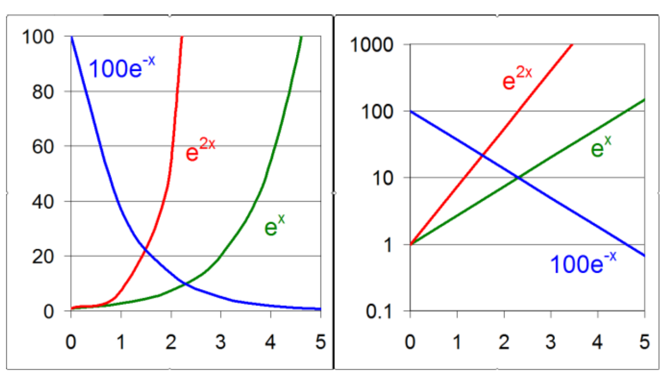


Figure 2-4 Exponential growth (red and green) and decay (blue). Left graph is linear. Right is semilog to fit both huge and tiny numbers.

Exponential growth and decay are mirror images. They may look like the right and left sides of parabolas, like the blue and green curves in the left panel of Figure 2-4 but they get much, much steeper, much, much faster than parabolas and their derivatives are never 0. At each step, exponential functions increase or decrease by the same ratio and soon become either astronomical or microscopic. In Figure 2-4, the red ($y = e^x$) and faster growing green ($y = e^{2x}$) curves grow exponentially while the blue curve ($y = 100e^{-x}$), which plots the history of my investments, decays exponentially. That makes me a good investor, but in negative time. My only comfort is that exponential decay never reaches zero while gambling will put you in debt.

Because exponential curves either quickly outgrow graphs with a linear scale or appear to shrink to 0, semilog graphs were created. In the right panel of Figure 2-4, x is linear but y increases exponentially with height. In a semilog graph, $y = e^{\pm kx}$ is a straight line with slope proportional to $\pm k$.

The Exponential as a Limit: A Taste of Differential Equations

Eqn. 2-10 is a *Differential Equation* because it contains derivatives. Differential equations may be easy to write, but are not so easy to solve. (Chapter 5 is dedicated to solving differential equations.) If bankers were able to solve differential equations they would not need to devote their lives to confiscating your money. Nevertheless, even if they don't know it, bankers use differential equations when they rip you off (i. e. give you a loan or a savings account). But that is another story. What we need now is to find that elusive exponential function, e^{kt} and prove that Eqn. 2-11 is the solution to Eqn. 2-10.

Here is a simple example to introduce exponentials and keep your interest at the same time. It is based on an idea called continuous compounding. At the end of year 0 you put $y(0)$ = \$1000 in the bank. The bank offers to pay you simple, annual interest at rate, k = 12% = 0.12 (dream on). Therefore at the end of year 1 you earn 12% of \$1000, or \$120. You then have a total of $y(1)$ = \$1120 at the end of year 1. You can write this as,

$$y(1) = y(0) \times (1 + k)$$

In other words, $y(1)$, the principal at the end of 1 year, equals $y(0)$, the starting principal, times $(1 + k)$, one plus the interest rate.

Now here is the way to outsmart the bank (fat chance). Instead of keeping your money in the full year, take it out after half a year. The bank should give you half the annual interest, or 6% of \$1000. This means that after half a year you would have \$1060. Then, open a new account, putting your money right back into the bank and keep it there for the next half year. The bank should pay you 6% of \$1060 for the second half of the year. This would produce 6% of \$1060 or \$63.60. Thus, by the end of the year you would have \$1123.60, an extra \$3.60 due to the fact that your interest has itself earned some interest for half a year! This is called compound interest and its mathematical form is,

$$y_{C2}(1) = y(0) \times \left(1 + \frac{k}{2}\right) \times \left(1 + \frac{k}{2}\right) = y(0) \times \left(1 + \frac{k}{2}\right)^2 = y(0) \times \left(1 + k + \frac{k^2}{2}\right)$$

Here, the subscript, $_{C2}$ indicates Compounding 2× per year.

Compounding is a very old bank trick. Just look at any loan. The more frequently the bank compounds, the more money it will make on loans. Banks compound loans as often as the law will allow them to get away with it (usually once a month).

In Calculus we look at limits. What we want to consider here is the most you can make in the limit of compounding continuously (every instant). So let's start simply. With compounding only once at the end of each year we already know that,

$$y_{C1}(1) = y(0) \times (1 + k)$$

Compounding twice a year yields

$$y_{C2}(1) = y(0) \times \left(1 + \frac{1}{2}k\right)^2$$

after a year. Continue the process by induction to compound *n* times per year. This yields,

$$y_{Cn}(1) = y(0) \times \left(1 + \frac{1}{n}k\right)^n$$

How much good will compounding more frequently do? We already know that simple interest yields \$1120 while compounding twice a year yields \$112<u>3.60</u>. Compounding three, four and five times a year yields \$112<u>4.86</u>, \$112<u>5.51</u>, and \$112<u>5.90</u> respectively. Compounding an infinite number of times - called ***continuous compounding***, yields \$112<u>7.50</u>. This is not much of an improvement, but remember: every penny you siphon away from the bank is a penny earned and a good deed done for humanity!

When we take the limit as $n \to \infty$ and compound continuously, we get the exponential,

$$y_{c\infty}(1) = y(0) \times \lim_{n\to\infty}\left(1 + \frac{1}{n}k\right)^n = y(0)e^k \qquad \text{Eqn. 2-13}$$

Hidden in this expression is the number, *e*, which we finally define as

$$\boxed{e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n} \qquad \text{Eqn. 2-14}$$

To get some idea of the limit as $n \to \infty$, lets calculate $(1 + 1/n)^n$. for a few values of *n*.

$$\left(1+\frac{1}{1}\right)^1 = 2 \cdots \left(1+\frac{1}{2}\right)^2 = 2.25 \cdots \left(1+\frac{1}{3}\right)^3 \approx 2.37 \cdots \left(1+\frac{1}{10}\right)^{10} \approx 2.5937$$

Clearly, we approach the limiting value, e ≈ 2.71828….very slowly.

Do you see how strange limits can be? If we took the limit of 1 + 1/*n* as $n \to \infty$ first we would simply get 1, and 1 raised to any power is 1, not 2.718….

Next, how do we handle *k* in the parenthesis of Eqn. 2-13? We can get it out of the parenthesis (and into the exponent) if we can prove that as $n \to \infty$,

$$\lim_{n\to\infty}\left(1 + \frac{1}{n}k\right)^n = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^{kn} = e^k \qquad \text{Eqn. 2-15}$$

This involves a bit of trickery. The way to get *k* out of the parenthesis is to set *k*/*n* equal to another number, namely *k*/*n* = 1/*m* (*km* = *n*). Since $n \to \infty$ so does $m \to \infty$ (smaller infinity is still infinity), and the two expressions are equal. Thus e^{kn} is given by Eqn. 2-16.

$$e^{kn} = \lim_{n\to\infty}\left(1+\frac{k}{n}\right)^{n} = \lim_{m\to\infty}\left(1+\frac{1}{m}\right)^{km} \qquad \text{Eqn. 2-16}$$

We must generalize further because few people keep money in the bank exactly one year. Many need it to pay off some gambling or credit card debt long before the year has expired. Of course, there are also dull people (who don't gamble their money away) who keep their money in the bank much longer than a year.

The problem is to compute how much money you will have at any given time, *t*. If interest is only paid once at the end of time, *t* then you will have simply,

$$y(t) = y(0)(1+kt)$$

If the interest is compounded *n* times in time period, *t*, then you will have,

$$y(t) = y(0)\left(1+\frac{k}{n}t\right)^{n}$$

And, if it is compounded continuously ($n \to \infty$) after time, *t* you will have,

$$y(t) = y(0)e^{kt} \qquad \text{Again, Eqn. 2-11}$$

In terms of Calculus, continuous compounding means that at every instant the rate of earning is proportional to the principal at that instant. Thus, I think we have proven that Eqn. 2-11 is the solution to the differential equation, Eqn. 2-10.

Eqn. 2-11 is designed to tell you how much money you will have after a specified time interval, *t*. If you want to know how long it will take to become a millionaire, then you must solve for time, *t.* You do this by taking the inverse exponential or logarithm of each side of the equation. Of course, it is natural that you have forgotten everything you learned about natural logs, or any logs, so the brief review in the next section may save your life.

2.4 Natural Logarithmortis

You can tell by the heading that the subject of logarithms (better known as *logs*) is going to be a real killer. Yet astoundingly, logarithms were inspired by sheer laziness. Logarithms replace the difficult processes of multiplying and dividing with the simpler processes of adding and subtracting. I'll bet you never knew that the word **log*arithm*** contains parts composed of parts of the words, **log**ic and ***arithm***etic. I didn't!

Logs are basically powers or exponents of numbers

$$\text{number} = \text{base}^{\log} = \text{base}^{\text{power}} \qquad \text{Eqn. 2-17}$$

Using n for number, p for power, log, or exponent, and b for base, gives Eqn. 2-17 a make-over as,

$$\boxed{n = b^p} \quad \leftrightarrow \quad \boxed{p = \log_b(n)} \qquad \text{Eqn. 2-18}$$

Eqn. 2-18 shows that logs and exponentials are inverse functions.

Logs start hard and branch out into greater difficulties. For example, we have learned to accept that 8 = 2^3, but find it much harder to accept that 7 ≈ $2^{2.807}$ because we have difficulty visualizing fractional powers. But though fractional powers are harder, they are just as valid. Here are a few cases of logs.

Hard:	When $5^2 = 25$,	2 is the log (power) of 25 to the base 5.
Harder:	When $5 = \sqrt{25} \equiv 25^{½}$,	½ is the log (power) of 5 to the base 25.
Hardest:	When $7 = 25^{0.6045}$,	0.6045 is the log (power) of 7 to the base 25.

What makes logs so useful? Instead of multiplying numbers add their logs. This might be called log arithmetic!

Hard Example: Multiply powers of equal numbers such as 2^3 by 2^2.
Hard Solution: Add their powers.

$$2^3 \times 2^2 = 2^{3+2} = 2^5$$

Harder Example: Multiply two unequal numbers such as 2 and 4 that seem related.
Harder Solution: A math talent might recognize that $4 = 2^2$ so that the problem becomes,

$$2 \times 4 = 2^1 \times 2^2 = 2^3$$

Almost Hardest Example: Multiply two unequal numbers such as 2 and 3, that don't seem related!
Almost Hardest Solution: A genius such as Napier realized not only that is $2 = 2^1$, but $3 \approx 2^{1.58496}$. Therefore,

$$2 \times 3 \approx 2^1 \times 2^{1.58496} = 2^{2.58496} \approx 5.99999$$

That's weird! I thought 2×3 = 6. The small error introduced by using logs is due to rounding. This is one of the problems of using logs in place of simple numbers.

The principle of multiplying numbers by adding their logs (or powers or exponents) led to the invention of the *slide rule*, which I and other suckers like me had to use because in the geologic epoch I took math, a scientific calculator was about as large as a room.

Getting Down to Bases

Here is a tough question. How do you decide which base to use? We can use any base we want. In fact, we could get totally ridiculous and multiply (2)(3) by writing it as $(1.75)^{1.2386}(1.75)^{1.9632}$.

What is the *best* base to use? In math courses Before Calculus (BC), base 10 was the most popular because, as you may remember, most of us have 10 fingers and 10 toes. However, in most cases the best base to use is e, because it pays to be natural!

Logs to the base *e* are called *natural* logs, and they have a special abbreviation.

Natural Logs and e

The log to the base *e* of a number, *n* is called the natural log, and is written as ln(*n*). If we set base *b* = e, Eqn. 2-18 becomes,

$$\boxed{n = e^{p} \quad \leftrightarrow \quad p = \log_e(n) \equiv \ln(n)} \qquad \text{Eqn. 2-19}$$

This makes natural logs and ***e*** inverse functions. Then, Eqn. 2-11 (the solution to Eqn. 2-10) can be written in natural log form in two ways (because ln[*a*/*b*] = ln[*a*] – ln[*b*]) as

$$\ln\left[\frac{y(t)}{y(0)}\right] = kt \quad \Leftrightarrow \quad \ln[y(t)] = \ln[y(0)] + kt \qquad \text{Eqn. 2-20}$$

Differentiating each side means that

$$\mathrm{d}\ln[y(t)] = k\mathrm{d}t \qquad \text{Eqn. 2-21}$$

Rearranging Eqn. 2-10 to solve for *kdt* yields

$$\frac{\mathrm{d}y}{y} = k\mathrm{d}t \qquad \text{Eqn. 2-22}$$

Eliminating *kdt* by equating Eqn. 2-21 and 2-22 yields

$$\boxed{\frac{\mathrm{d}y}{y} = \mathrm{d}\ln(y)} \qquad \text{Eqn. 2-23} \qquad\qquad \boxed{\frac{\mathrm{d}\ln(y)}{\mathrm{d}y} = \frac{1}{y}} \qquad \text{Eqn. 2-24}$$

Thus, the derivative of ln*(y)* is *1/y*!

Since the integral is the reverse of the derivative, integrating Eqn. 2-24 shows that the *integral* of *1/y* with respect to *y* is ln(*y*)! Backing into integrals is so much fun. I can tell you that I was shocked the first time I saw this result. The function, *1/y* seems so simple I never would have guessed that its integral is as complicated as ln(*y*). But in retrospect, what could be more natural? In fact, you will see in the next chapter why the natural log is usually defined as the integral,

$$\boxed{\ln(x) \equiv \int_1^{x} \frac{\mathrm{d}y}{y}} \qquad \text{Eqn. 2-25}$$

Here, *x* only appears as the upper limit of the integral and *y* is the variable of the integrand of Eqn. 2-25. It is called a dummy variable because it doesn't appear outside the definite integral, so it doesn't matter what symbol is used inside the integral. I could have used any symbol. If I had used *cabin*, it would be the standard joke, "What is the integral of d*cabin* over *cabin*?" Answer: …natural log *cabin* – but only if the integral's upper limit is *cabin*!

2.5 Exponentials and Logs: Spirals and the Age of the Earth

We now use exponentials and natural logs to determine the Age of ancient creatures such as the dinosaurs, and of the Earth itself. Beginning around 1800 geologists realized that so many sedimentary rock layers are piled one on top of another that to accumulate these layers the Earth must be very ancient indeed. They recognized that the more geologically recent rock layers (for at least the past 540 million years) contain key fossils showing that many species have appeared, dominated the Earth for a time, and then disappeared. They recognized that this long chain must have taken an immense time to form, but for over a century the geologists could not put a number to the age of the Earth. James Hutton even used the now famous quote, "No vestige of a beginning: No prospect of an end".

For years, Lord Kelvin, a prodigious physicist and mathematician, scorned the geologists' qualitative impression that the Earth was at least hundreds of millions of years old, silencing them with mathematical arguments and his reputation as a genius. This was just one more instance where an arrogant mathematician has been consummately wrong. But it is a clear case of don't blame the mathematics; blame the mathematician (and the timid geologists).

Around 1900, Ernest Rutherford and Frederick Soddy discovered that radioactive elements decay at fixed rates and soon realized it gave them equations to determine the age of rocks and ultimately, of the Earth. Many minerals, such as zircons, crystallize with radioactive elements in their matrix. With time, the number, $N(t)$ of radioactive isotopes decreases from the original number, $N(0)$ at a rate that is proportional to $N(t)$. Thus, the governing equation is Eqn. 2-10, with $N(t)$ in place of $y(t)$ and $-k$ instead of k.

$$\boxed{\frac{\mathrm{d}N(t)}{\mathrm{d}t} = -kN(t)} \qquad \text{Eqn. 2-26}$$

Using Eqn. 2-20 as the solution of Eqn. 2-26, the age, t is,

$$\boxed{t = -\frac{1}{k}\ln\left(\frac{N(t)}{N(0)}\right)} \qquad \text{Eqn. 2-27}$$

Geologists add one further twist because they prefer to think in terms of the half life, $t_{1/2}$, the time needed for the number of radioactive atoms to decrease by 50%.

$$t_{1/2} = -\frac{1}{k}\ln\left(\frac{1}{2}\right) = +\frac{1}{k}\ln(2) \qquad \text{Eqn. 2-28}$$

Then the solution for age, t (Eqn. 2-28) becomes,

$$\boxed{t = \frac{t_{1/2}}{\ln(2)}\ln\left[\frac{N(0)}{N(t)}\right]} \qquad \text{Eqn. 2-29}$$

Problem: If a zircon crystal originally contained 91000 U-235 atoms but now contains only

1000, calculate its age.
Information: Uranium 235 has a half life of just over 700 million years, so $t_{1/2}$, ≈ $7(10)^8$.
Solution: All we have to do is substitute in Eqn. 2-29

$$t_{\text{EARTH}} = \frac{7(10)^8}{\ln(2)} \ln\left[\frac{91000}{1000}\right] \approx 4.56(10)^9$$

Problem: If I earn y(0) = $1,000 from this book and put it in the bank today, calculate the number of years, *t* it will take to grow to $1,000,000 at *k* = 1% interest with continuous compounding.
Solution: Since *k* is a growth constant in this problem and not a decay constant, the solution is like Eqn. 2-27, but with a positive sign on the right hand. Thus

$$t = \frac{1}{k}\ln\left[\frac{N(t)}{N(0)}\right] = \frac{1}{.01}\ln\left[\frac{10^6}{10^3}\right] \approx 690.77$$

According to this calculation, by the time I become a millionaire, the age of the Earth will have almost doubled. (Alright, I am exaggerating a bit, but the point is that I won't be around to collect. On second thought, it might have been better to be a banker, after all!)

If we solve for *N*(*t*) the number of radioactive isotopes (or the money) at any given time, *t*, take the inverse natural log of each side of Eqn. 2-27. This yields,

$$\boxed{N(t) = N(0)e^{-kt}} \qquad \text{Eqn. 2-30}$$

This is identical to Eqn. 2-11. Whenever we describe decay, we must either make the exponent negative or make *k* negative. In this book, I have chosen to make the exponent negative for decay problems and make *k* a positive decay constant.

Problem: The Dutch paid the Canarsie Indians *N*(0) = $24 for Manhattan in 1624. The Indians invested it in the Dutch stock market at 7%. How much are they worth now?
Solution: It is now roughly 390 years later so *t* = 390 and *k* = 0.07. Use Eqn. 2-30 with a plus sign in the exponent since *k* is a growth rate.

$$N(400) = \$24e^{0.07(390)} \approx \$24e^{27.3} \approx \$1.72(10)^{13}$$

In everyday lingo this is 17.2 trillion dollars. Why, it would settle the National debt and then some! The Canarsie Indians would own a good part of the world! And the irony is that they cheated the Dutch because they snuck over from Brooklyn to Manhattan!

The Exponential Function, a^x: Combining Exponentials and Logs

What is the derivative of a^x with respect to *x* when *a* is a constant, $a \neq \boldsymbol{e}$? Eqn. 2-19 shows that the relation between *a* and ***e*** is,

$$a = e^k \quad \Leftrightarrow \quad \ln(a) = k$$

Then

$$\boxed{\frac{\mathrm{d}a^x}{\mathrm{d}x} = \frac{\mathrm{d}e^{kx}}{\mathrm{d}x} = ke^{kx} = a^x \ln(a)}$$ Eqn. 2-31

Wrapping Up Exponentials: The Logarithmic Spiral

I spent a great amount of time thinking about how to wrap up this section on exponentials and logs. Then I realized that Nature has found a compact way to wrap up exponentials into neat packages called logarithmic spirals.

A circle is a curve where the radius remains the same as the angle increases. A spiral is a curve where the radius increases as the angle increases. When you wrap a garden hose the radius of the spiral increases linearly with angle and the space between successive turns is constant. When a nautilus builds its shell as it grows, it needs more space, so the space between successive turns gets larger as the angle increases. This is what gives birth to the logarithmic spiral.

This book started with a photo of the logarithmic spiral of a nautilus shell. Figure 2-5 suggests that the logarithmic spiral is a form taken by many objects in nature from Romanesco (spiral) broccoli to cloud patterns of intense winter storms to spiral galaxies.

Figure 2-5 Romanesco broccoli (left), cyclonic storms (center) and spiral galaxies (right) often match logarithmic spirals.

The equation for the logarithmic spiral states that radius increases exponentially with angle starting from an initial radius r_0 that is not zero.

$$\boxed{r = r_0 e^{k\theta}}$$ Eqn. 2-32

To get the logarithmic form, take the natural log of each side and solve for the angle, θ.

$$\boxed{\theta = k \ln\left[\frac{r}{r_0}\right]}$$ Eqn. 2-33

Figure 2-6 shows two logarithmic spirals that curl in opposite directions. Nature sometimes appears to select the sense of rotation carefully and does not seem to care at other times. Isn't nature interesting?

A helix is a spiral or circle that rises into the 3rd dimension. It is a curve where height and possibly radius, increase as the angle increases. A spiral staircase is really a helix with a constant radius. So is DNA, the genetic code of life, when it unwinds to form a double helix. Death may come from violent swirling winds that sometimes widen on top to resemble a logarithmic helix called a tornado (Fig. 2-7). The helix shape can create great strength out of weakness. Twisting fragile cotton fibers forms an intertwined helix called thread and whose strength is far, far greater than the sum of the individual fibers. The strong connective tissue called collagen is a helix formed of twisted fibrils of protein. Even the heart exhibits something of a spiral or helical shape.

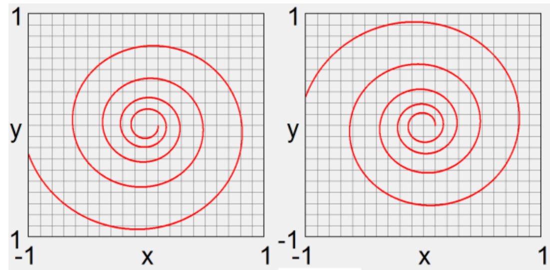


Figure 2-6 Logarithmic spirals with opposite senses of rotation.

Figure 2-7 Helical winds may form a tornado.

Equations for a helix in which height increases linearly with angle, θ, are,

$$\boxed{\begin{aligned} x &= r\cos(\theta) \\ y &= r\sin(\theta) \\ z &= k\theta \end{aligned}} \qquad \text{Eqn. 2-34}$$

Radius, *r* may stay constant, or may increase linearly or exponentially with angle, θ.

2.6 Derivative of the Inverse Sine

Sometimes you are given a sine from heaven and can't quite believe it. If so you will surely ask, "What's the angle?" Remember that just as the log has an inverse, so do the trig functions. The *inverse sine* is simply the angle that corresponds to a given sine.

Finding the derivative of the $\sin^{-1}(x)$ is extremely tricky because it requires 5 easy pieces (or steps. The math is easier than that screwed up movie). Take the sine of each side, differentiate, use the trig identity to replace cosine with sine

$$y = \sin^{-1}(x)$$

Step 1 $x = \sin(y)$

Step 2 $\mathrm{d}x = \cos(y)\mathrm{d}y$

Step 3 $\mathrm{d}x = \sqrt{\left[1 - \sin^2(y)\right]}\mathrm{d}y$

Step 4 $\mathrm{d}x = \sqrt{\left[1 - x^2\right]}\mathrm{d}\left[\sin^{-1}(x)\right]$

Finally, with **Step 5** we rearrange to arrive at the solution

$$\boxed{\frac{\mathrm{d}\left(\sin^{-1}(x)\right)}{\mathrm{d}x} = \frac{1}{\sqrt{1-x^2}}} \qquad \text{Eqn. 2-35}$$

Why don't you try to find the derivative of the inverse cosine. Don't be a chicken! The procedure is the same but the result has a minus sign, as you might have expected,

$$\boxed{\frac{\mathrm{d}\left[\cos^{-1}(x)\right]}{\mathrm{d}x} = -\frac{1}{\sqrt{1-x^2}}} \qquad \text{Eqn. 2-36}$$

Finally, the derivative of $tan^{-1}(x)$ with respect to x is,

$$\boxed{\frac{\mathrm{d}\left[\tan^{-1}(x)\right]}{\mathrm{d}x} = \frac{1}{1+x^2}} \qquad \text{Eqn. 2-37}$$

(Notice the + sign before x^2 in the radical.)

Now it's time to back into another integral. If someone had told you to integrate,

$$\int_{x=0}^{x} \frac{\mathrm{d}x}{\sqrt{1-x^2}}$$

before you saw Eqn. 2-35 you would not have had a gnat's change in a spider web to realize it is,

$$\boxed{\sin^{-1}(x) = \int_{x=0}^{x} \frac{\mathrm{d}x}{\sqrt{1-x^2}}} \qquad \text{Eqn. 2-38}$$

Isn't it wonderful that any function is the integral of its derivative! It enables us to work backwards instead of working in the dark. Working backwards is so much fun! Why don't you try backing into the integral,

$$\int_{x=0}^{x} \frac{\mathrm{d}x}{1+x^2} = ?$$

(Hint: Use Eqn. 2-37.) Oh, are we getting smart! Or am I going off on an inverse tangent?

2.7 Chain Gang Rules, Etc. – Derivatives of Complicated Functions

Just imagine you are basking in glory on a tropical island, proud that you know the derivatives of such advanced functions as sin(x), e^x, ln(x), and $\sin^{-1}(x)$ while the mass of the human race couldn't even have a hint of your elevated thoughts. **Warning:** Suddenly, a math nerd comes shuffling down the beach to spoil all your fun. "What", asks the nerd with a smirk, "is the derivative of $e^x \cdot \sin(x^2)$?" What can you do? You've never seen such a complicated function before. Whatever you do, don't panic. Stop, take a long breath, and calm yourself by recalling that even Newton screwed up the first time he tried to find the derivative of the product of two functions. Once you've regained your composure, tell the nerd you'll respond at tonight's beach party. Not only will this put you on the offensive, it will give you time to read the rest of this section.

The purpose of this section is to reveal the rules for finding the derivative of any complicated function such as a sum, product, power, or function of simpler functions. The rules are much like the trick for pronouncing complicated words. When I was in the first grade I could only read words with one syllable. In the second grade I suddenly realized (or was told) that even the longest words such as *bron·to·sau·rus* were easy to read because they consisted of short syllables that I could pronounce and string together. You can find the derivative of any complicated function so long as you know the derivative of each of its parts. For example, the derivatives of various combinations of x^2 and $sin(x)$ are given in Table 2-1 But what are the general formulas that produced these derivatives? For this we need…

	Function	Derivative
Sum	$x^2+\sin(x)$	$2x+\cos(x)$
Product	$x^2\sin(x)$	$x^2\cos(x)+2x\sin(x)$
Quotient	$\dfrac{1}{\sin(x)}$	$-\dfrac{\cos(x)}{\sin^2(x)}$
Fraction	$\dfrac{x^2}{\sin(x)}$	$\dfrac{2x\sin(x)-x^2\cos(x)}{\sin^2(x)}$
Fct of Fct	$\sin(x)$	$2x\cos(x^2)$

Table 2-1 Derivatives of functions that are sums, products, quotients, fractions or functions of functions.

The Syllables – u and v

In Calculus, the letters, u and v often indicate the simple parts (syllables) of complicated functions. As an example, if the function is $x^2 sin(x)$ the syllables are $u = x^2$ and $v = sin(x)$. The derivatives of the various combinations of simple functions, u and v are given in Table 2-2,

Function	Derivative
$u+v$	$\dfrac{du}{dx}+\dfrac{dv}{dx}$
uv	$u\dfrac{dv}{dx}+v\dfrac{du}{dx}$
$\dfrac{1}{u}$	$-\dfrac{\frac{du}{dx}}{u^2}$
$\dfrac{u}{v}$	$\dfrac{v\frac{du}{dx}-u\frac{dv}{dx}}{v^2}$
$v(u)$	$\dfrac{dv}{du}\dfrac{du}{dx}$

Table 2-2 Rules for derivatives of complex functions.

Let's prove these one by one. To do so you must remember that the derivative is the limit of the ratio of differences.

The Addition Rule for Derivatives

The addition rule is so simple I am almost embarrassed to include it. It is merely a rearrangement. Thus, by definition

$$\frac{\mathrm{d}[u+v]}{\mathrm{d}x}=\lim_{\Delta x\to 0}\frac{[u+\Delta u+v+\Delta v]-[u+v]}{\Delta x}\qquad\text{Eqn. 2-39}$$

Simply cancel and rearrange terms and,

$$\frac{\mathrm{d}[u+v]}{\mathrm{d}x}=\lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x}+\lim_{\Delta x\to 0}\frac{\Delta v}{\Delta x}=\frac{\mathrm{d}u}{\mathrm{d}x}+\frac{\mathrm{d}v}{\mathrm{d}x}$$

Voila (even if you don't speak French), the addition rule itself!

The Product Rule for Derivatives

Are you still with me? The product rule is a wee bit tougher to prove. The definition,

$$\frac{\mathrm{d}[uv]}{\mathrm{d}x}=\lim_{\Delta x\to 0}\left[\frac{[(u+\Delta u)\times(v+\Delta v)]-[uv]}{\Delta x}\right]$$

followed by cancellation and rearrangement

$$\frac{\mathrm{d}[uv]}{\mathrm{d}x}=\lim_{\Delta x\to 0}\frac{[u\Delta v+v\Delta u+\Delta u\Delta v)]}{\Delta x}=\lim_{\Delta x\to 0}\left[u\frac{\Delta v}{\Delta x}\right]+\lim_{\Delta x\to 0}\left[v\frac{\Delta u}{\Delta x}\right]+\lim_{\Delta x\to 0}\left[\frac{\Delta u\Delta v}{\Delta x}\right]$$

gets us to

$$\frac{\mathrm{d}[uv]}{\mathrm{d}x}=u\frac{\mathrm{d}v}{\mathrm{d}x}+v\frac{\mathrm{d}u}{\mathrm{d}x}+\lim_{\Delta x\to 0}\left[\frac{\Delta u\Delta v}{\Delta x}\right]$$

This would be perfect if we could only get rid of the last term on the right. Once again, just as we seem to approach a final result in Calculus another insuperable barrier rears its ugly head. In fact, if I didn't know the answer, I might just have reached my limit.

As a matter of fact, taking a limit is precisely how to get rid of the extra little piece,

$$\lim_{\Delta x\to 0}\left[\frac{\Delta u\Delta v}{\Delta x}\right]$$

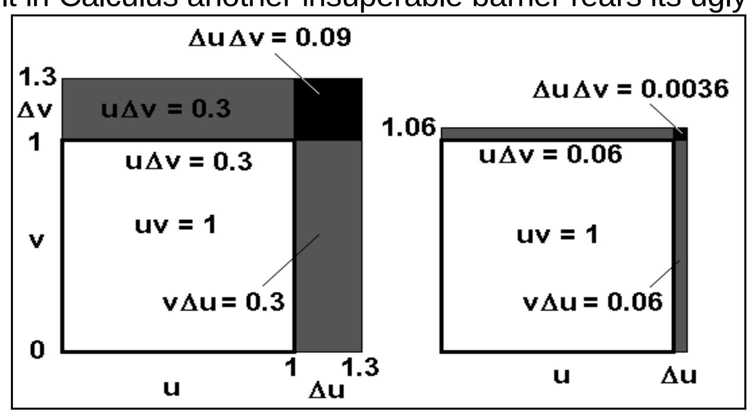


Figure 2-8 Why $\Delta u{\cdot}\Delta v$ is negligible as Δu and $\Delta v \to 0$

The denominator is tiny but the numerator is the product of two tiny terms, which makes it very, very tiny. Figure 2-8, which we have seen before in another life (Figure 1-10) shows that as *Δu* and *Δv* get small, the product, *ΔuΔv* gets much smaller than either, (e. g., 0.003×0.002 =

0.000006) and in the limit can be neglected. So, finally we have,

***The product rule*,**

$$\boxed{\frac{\mathrm{d}[uv]}{\mathrm{d}x} = u\frac{\mathrm{d}v}{\mathrm{d}x} + v\frac{\mathrm{d}u}{\mathrm{d}x}} \qquad \text{Eqn. 2-40}$$

The product rule is amazingly useful and versatile. (It forms the basis of Integration by Parts – see Section 3.3.) Let's try it on a function whose derivative we know. We know that the derivative of $y = x^3$ is $dy/dx = 3x^2$. Now, split x^3 artificially to $u{\cdot}v = x{\cdot}x^2$. Then,

$$\frac{\mathrm{d}[uv]}{\mathrm{d}x} = x\frac{\mathrm{d}[x^2]}{\mathrm{d}x} + x^2\frac{\mathrm{d}x}{\mathrm{d}x} = x\cdot 2x + x^2\cdot 1 = 3x^2$$

That works! Next, use the product rule to find the derivative of $1/u$ in terms of du/dx. To do this, set $u{\cdot}v = 1$ and therefore $1/u = v$. Then, the derivative of each side is,

$$uv = 1 \quad \Leftrightarrow \quad \frac{1}{u} = \mathrm{v} \quad \Leftrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1}{u}\right] = \frac{\mathrm{d}v}{\mathrm{d}x}$$

Here is the trick. Take the derivative of u·v. Why? When $1/u = v$ then $uv = 1$, so $\mathrm{d}[u{\cdot}v]/\mathrm{d}x = 0$. Then, use the equation directly above and replace v with $1/u$ to get

$$\frac{\mathrm{d}[uv]}{\mathrm{d}x} = 0 = u\frac{\mathrm{d}v}{\mathrm{d}x} + v\frac{\mathrm{d}u}{\mathrm{d}x} = u\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1}{u}\right] + \frac{1}{u}\frac{\mathrm{d}u}{\mathrm{d}x}$$

When you solve this for $\mathrm{d}/\mathrm{d}x[1/u]$ it yields,

The Inverse rule

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1}{u}\right] = -\frac{1}{u^2}\frac{\mathrm{d}u}{\mathrm{d}x}} \qquad \text{Eqn. 2-41}$$

This result will also be extremely useful. In fact, if I wanted to, I could now easily combine the product rule and the inverse rule to get

The quotient rule

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{u}{v}\right] = \frac{1}{v^2}\left(v\frac{\mathrm{d}u}{\mathrm{d}x} - u\frac{\mathrm{d}v}{\mathrm{d}x}\right)} \qquad \text{Eqn. 2-42}$$

There it is! However, right now I am feeling particularly lazy, so I'll let you derive it. And while you are straining your brain I'll use the quotient rule to solve a simple but important...

Problem: Find the derivative of tan(x).
Solution: Since tan(x) = sin(x)/cos(x), its derivative is tailor made for the quotient rule.

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\sin(x)}{\cos(x)}\right]=\frac{1}{\cos^2(x)}\left(\cos(x)\frac{\mathrm{d}[\sin(x)]}{\mathrm{d}x}-\sin(x)\frac{\mathrm{d}[\cos(x)]}{\mathrm{d}x}\right)=\frac{\cos(x)\cos(x)-\sin(x)[-\sin(x)]}{\cos^2(x)}$$

Since the numerator is simply [$\cos^2(x)$ + $\sin^2(x)$] = 1, d[tan(x)]/dx reduces to 1/$\cos^2$(x), which, using trigonometric identities, can be expressed in the following ways.

$$\boxed{\frac{\mathrm{d}[\tan(x)]}{\mathrm{d}x}=\frac{1}{\cos^2(x)}=\sec^2(x)=\tan^2(x)+1} \qquad \text{Eqn. 2-43}$$

The Chain Rule

The chain rule is simply a product. Breaking down the function u (which is a function of v, which, in turn, is a function of x) into its syllables, the chain rule is,

$$\boxed{\frac{\mathrm{d}u[v(x)]}{\mathrm{d}x}=\frac{\mathrm{d}u}{\mathrm{d}v}\frac{\mathrm{d}v}{\mathrm{d}x}} \qquad \text{Eqn. 2-44}$$

I will now use the chain rule to find the derivative of complicated functions such as sin(x^2) and $(x^2 + 3x)^4$. So, with u = sin(v) and $v = x^2$, dsin(x^2)/dx is,

$$\frac{\mathrm{d}\left[\sin\left(x^2\right)\right]}{\mathrm{d}x}=\frac{\mathrm{d}[\sin(x^2)]}{\mathrm{d}[x^2]}\frac{\mathrm{d}[x^2]}{\mathrm{d}x}=\cos(x^2)\cdot 2x$$

And, with $u = v^4$ and $v = (x^2+3x)$, and d[$(x^2+3x)^4$]/dx is,

$$\frac{\mathrm{d}\left[\left(x^2+3x\right)^4\right]}{\mathrm{d}x}=\frac{\mathrm{d}\left[\left(x^2+3x\right)^4\right]}{\mathrm{d}\left[\left(x^2+3x\right)\right]}\frac{\mathrm{d}\left[\left(x^2+3x\right)\right]}{\mathrm{d}x}=4\left(x^2+3x\right)^3\cdot\left(2x+3\right)$$

(Brief) Triumph

Now, at long last you can take all sorts of derivatives, so you can take on the math nerd. True, you may never know all the math that a math nerd knows, but you are a lot sexier.

Chain Rule and Evaporating Hopes

Any mathematician can dream up complicated functions that need the chain rule. But when would you encounter the chain rule in the real world? Imagine that you are dying of thirst in a desert where the relative humidity, RH = 20%. Rain begins to fall from the clouds 1 mile up. You stick your tongue out to catch the falling raindrops. Unfortunately, the drops evaporate as they fall through the dry desert air. Assume the raindrops 1 mile up have a radius, r = 0.00002 m when they fall out of the cloud's base. Will they fall far enough to reach your tongue? Only the chain rule can tell.

You need an expression for dr/dz, but the laws are only given for dr/dt, the evaporation rate and dz/dt = U_t, the fall speed or terminal velocity of the drops. Both depend on the radius, r, of the drops.

$$\frac{\mathrm{d}r}{\mathrm{d}t} \approx 1.5(10)^{-10}\frac{(RH-1)}{r} \qquad \frac{\mathrm{dz}}{\mathrm{d}t} \equiv U_t \approx 1.25(10)^8 r^2 \quad \Leftrightarrow \quad \frac{\mathrm{d}t}{\mathrm{dz}} = \frac{1}{U_t}$$

This is a job for the chain rule because you must find d*r*/d*z* to determine how far the drops can fall before evaporating completely.

$$\frac{\mathrm{d}r}{\mathrm{dz}} = \frac{\mathrm{d}r}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{dz}} = \frac{1}{U_t}\frac{\mathrm{d}r}{\mathrm{d}t} \approx \frac{1.2(RH-1)(10)^{-18}}{r^3}$$

Oops! I'm sorry. This is a differential equation. To solve it, we must integrate. I will simply have to postpone the solution until the next chapter. But now that I must hold you in suspense, the least I can do now is show you how to use derivatives to...

2.8 Get the Most Out of Life Giving the Least: Maxima and Minima

Almost all of us would like to live with unlimited options and no problems. Only a few people actually relish problems. They are called mathematicians. Scientists, by contrast, relish solutions! (Or is it the other way around?) In any event, most of us hope to get the most out of everything good and suffer the least from everything bad.

Our job in life is to maximize the positives and minimize the negatives. Maximizing and minimizing always brings us back to derivatives.

The most I can do now is to show you how to find the maximum value of functions. These problems involve one simple trick that I will introduce by telling a story.

I grew up in Bayswater, in Far Rockaway, NY. About 4 miles to the west on the peninsula was Rockaway's Playland, which is now long gone, replaced by upscale, ground-level housing that in turn was replaced by Hurricane Sandy (2013). Only memories remain. I loved the Funhouse as a kid but wasn't allowed on the Roller Coaster. We of course heard that Coney Island's ***Cyclone*****,** shown in Figure 2-9, was ***<u>The</u>*** Roller Coaster, but that was far away. Finally, I got to go there, paid my money, and squeezed into a seat with my friend, Bob Salzman. The Cyclone started, slowly but steeply lifting me to my doom. At the top, the tracks leveled off. For an instant, the view was superlative.

Figure 2-9 Red line at the top of Coney Island's Cyclone Roller Coaster shows that at a Max the slope or derivative is 0. Min is out of sight on the left.

For a brief moment I thought of my Calculus. (I thought longer about my life.) I was at the top, the maximum height. I could see that the tracks were level, meaning the slope or derivative was zero, but the second derivative was negative, just like my prospects.

Just past the crest the tracks made a turn to the right that added to my terror. It looked like I was flying off into space. Down we went and up went my heart into my throat. I dared not swallow. At the bottom or the minimum height the tracks leveled off again. This meant the derivative was zero again, but this time the second derivative was positive. The coaster then jerked sharply upward. I tried to breathe but couldn't as the acceleration shoved my lungs down into my diaphragm. Bob was in ecstasy, while I wimped out. My only compensation was that I knew Calculus and he didn't. That meant he laughed at me for only five minutes. He begged me to go again but I told him I was reflecting on the nature of the second derivative.

So, what is the fundamental lesson here? When a function is

A: At its maximum the derivative is zero and the second derivative is negative.
B: At its minimum the derivative is zero and the second derivative is positive.

To prove this, go back to the definition of derivatives. At the top of the roller coaster the track must be lower both a bit in front and a bit behind. Using the forward difference, we see that if the derivative is positive then the point slightly ahead [y(x+Δx)] will be higher. In that case, y(x) cannot be a maximum. Using the backward difference, we see that if the derivative is negative then the point slightly behind [$y(x\text{-}\Delta x)$] was higher. In that case, $y(x)$ cannot be the maximum. By the process of elimination, if $y(x)$ is a maximum the derivative can only be 0.

Now let's show that the second derivative is negative at the top. It is best to use the centered difference formulation of the second derivative, Eqn. 2-9b.

$$\boxed{\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} \approx \frac{y(x+\Delta x)+y(x-\Delta x)-2y(x)}{\Delta x^2}}$$

At the maximum, the function is greater than at all points slightly behind or slightly ahead. Therefore, the second derivative must be negative because both [$y(x+\Delta x)$ - $y(x)$] < 0 and [$y(x$ - $\Delta x)$ - $y(x)$] < 0 are negative.

Case closed. Well, maybe the case isn't closed. Let's make the point visual. Figure 2-10 illustrates the same points as the reasoning above. You can see that at the maximum, the curve is horizontal. You can also see that at x - Δx the slope or derivative is positive but that it is negative at x + Δx. Thus the derivative of f(x) is decreasing at x, which means its derivative [the second derivative of f(x)] is negative. Using a picture for help may not seem impressive, but as is often the case, it is very effective.

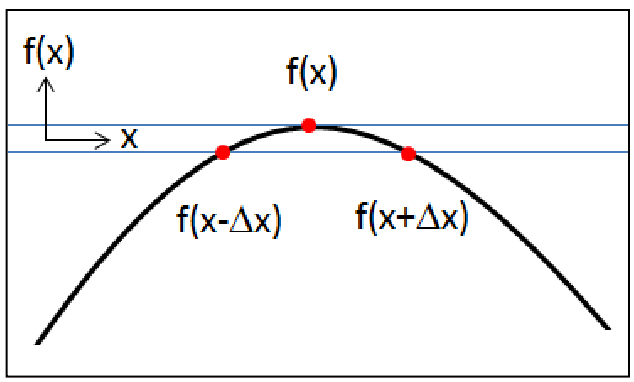


Figure 2-10 At a maximum, f(x) > f(x±Δx) and f'(x) = 0

Finding Extrema (Maxima and Minima)

Here is the technique for finding extrema.

1: Locate the extrema.

2: Find the values of the function at the extrema.

Extrema occur where the function's derivative is zero. Therefore,

To locate a max or min,

1: Take the derivative of the function,
2: Set it equal to zero.
3: Solve.

To find the value of the max or min,

1: Go back to the original equation,
2: Substitute the values for x_{ext}.
3: Solve.

Let's do this for a cubic equation, because a cubic is one of the simplest functions with both a local maximum and a local minimum. A local maximum is like a peak that may be surrounded by higher mountains, but is the highest point nearby.

Figure 2-11 The cubic equation, $y = (x - 1)(x - 5)(x - 6)$

Problem: Find the local max and min of the cubic equation, $y = (x - 1)(x - 5)(x - 6)$.

Hint: It always helps to plot the function. In Figure 2-11 I estimated just by looking at the graph that a local maximum occurs at $x \approx 2.5$, where $y_{max} \approx 12.5$, and a local minimum occurs at $x \approx 5.5$, where $y_{min} \approx -1.5$. Your eyeball estimates may be slightly different.

Solution: First, expand the cubic equation

$$y = (x-1)(x-5)(x-6) = x^3 - 12x^2 + 41x - 30$$

To locate the **extrema**, take the derivative and set it equal to 0. Then, $y = y_{ext}$ and $x = x_{ext}$.

$$\frac{\mathrm{d}y_{ext}}{\mathrm{d}x} = 3{x_{ext}}^2 - 24x_{ext} + 41 = 0$$

Do you know what to do now? Please don't tell me you still need help here! This is just algebra! (Don't you hate that expression?) You have a quadratic equation for x_{ext} so use the quadratic formula to find x_{ext}. This yields two values, namely,

$$x_{ext} = \frac{24 \pm \sqrt{(24)^2 - 4(3)(41)}}{6} = \frac{24 \pm \sqrt{84}}{6} \approx \begin{matrix} 5.53 \\ 2.47 \end{matrix}$$

Note that these values are very close to my estimates of x_{ext} from merely looking at Figure 2-11.

Now that we have the locations of the extrema, we must find their values. How in God's name do we do this? It's not so hard if you think about it. (Plug in x_{ext} to get y_{ext}!)

$$y_{ext} = {x_{ext}}^3 - 12{x_{ext}}^2 + 41x_{ext} - 30 = \begin{matrix} 5.53^3 - 12(5.53)^2 + 40(5.53) - 30 = 13.1 \\ 2.47^3 - 12(2.47)^2 + 40(2.47) - 30 = -1.1 \end{matrix}$$

Once again, my visual estimates (12.5, -1.5) from Figure 2-11 were reasonably close to the solutions. It always pays to graph a function. Seeing is a great, and often very underrated, guide.

Maximizing Profits

Permit me to be your profit maximizing prophet. Profit equals revenues minus costs, so the best way to proceed is to diagnose each revenue and cost factor separately, combine them by adding all revenues and subtracting all costs. Then maximize by taking the derivative and setting it equal to zero.

Consider a crude business model built of simple equations for each aspect of revenue and cost shown in Figure 2-12. Real businesses conduct more realistic analyses but the fundamental approach is the same.

Gross revenue is linearly proportional to N, the number of items sold. **Set it = $8×N** even though we know that the price cannot be maintained if we produce an infinite number of items. When production increases above demand the price per item must be reduced. Eventually, no more books can even be given away! After all, by the latest statistics, only about 1000 students take Calculus each year in the USA (not true) and there are about 1000 different Calculus books (almost true).

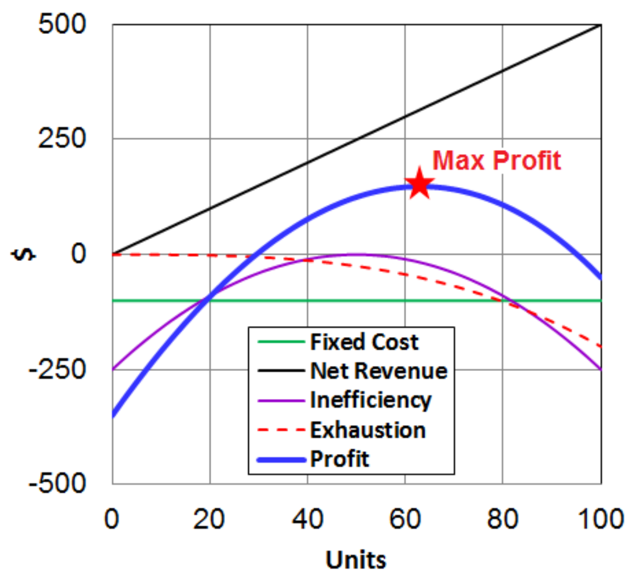


Figure 2-12 Financial analysis of this book.

Fixed costs include the building and machinery. **Set them = -$100.** (All costs are negative.) Variable costs include material and labor and depend almost linearly on the number, N of items produced. **Set them = -$3×*N***. There is inefficiency of scale, which is large when the business is either too small or too large. **Set it = -$0.1×(*N*-50)²**, the up-side-down purple parabola. Finally there is exhaustion of resources. If volume gets too high, we will run out of ink, trees, and paper. This cost would skyrocket. Therefore, **set it = -$0.0002×*N*³**, the red dashed cubic curve.

Summarizing, the simplified financial model consists of,

1: The constant term = Fixed Costs = -$100
2: The linear term = Gross Revenue - Variable Costs = $5×$N$
3: The quadratic term = Inefficiency = -$0.1($N$-50)^2
4: The cubic term = Exhaustion = -0.0002N^3$

Profit, P equals the revenues minus the costs. The equation for profit is therefore,

$$P = -0.0002\,N^3 - 0.1(N-50)^2 + 5N - 100 \qquad \text{Eqn. 2-45}$$

We find the maximum profit, P_{max} the new-fashioned (Calculus) way – Do you remember the steps? **Warning**: The number of items, N is the dependent variable.

1: Take the derivative of Eqn. 2-45 with respect to *N*.

$$\frac{\mathrm{d}P}{\mathrm{d}N} = -3\left[0.0002\,N^2\right] - 2\left[0.1(N-50)^1\right] + 5$$

2: Set the derivative equal to zero

$$0.0006N^2 + 0.2N - 15 = 0 \qquad \text{Eqn. 2-46}$$

3: Solve Eqn. 2-46 for *N* the old-fashioned (Algebra) way. Eqn. 2-46 is quadratic and has two solutions.

$$N = \frac{-0.2 \pm \sqrt{0.04 - 4(-15)(0.0006)}}{0.0012} = \begin{matrix} +63.06 \\ -396.4 \end{matrix} \qquad \text{Eqn. 2-47}$$

This steps locates *N*, the number of books that will produce the maximum profit.

4. Choose the positive value of *N* because it is impossible to sell a negative number of books, no matter how bad they are. Note that this matches the starred point in Figure 2-12.

5. Substitute the positive value of *N* = 63 (not 63.09 because no one can sell 0.09 books) from Eqn. 2-47 back into Eqn. 2-45 to find P_{max} = $148.09

Parting Words

I really hope this chapter has maximized your profit. Knowledge and understanding are the greatest of all profits. Now you should be ready to integrate your knowledge.

CHAPTER 3: INTEGRALS: DERIVATIVES IN REVERSE

Authors, teachers, and students debate about the order of topics in any complex subject. Some subjects, such as history, are natural to arrange chronologically because the past determines (and if we are lucky, guides) the future. In mathematics and science, it is also wise to arrange the topics chronologically if possible because there is often a natural order in which concepts and ideas were discovered. But I have been guided by a more profound principle, namely your short memory of the derivatives you just learned and will need to remember to do many integrals. But I am wasting time and every minute I waste you may forget another derivative. So before you forget every one, let's start.

3.1 Integrals are Antiderivatives: The Fundamental Theorem of Calculus

It is supremely important to remember derivatives when doing integrals because integrals are just derivatives in reverse. (Are you getting tired of reading this?)

Surely, you remember Eqn. 1-23, namely that the derivative of x^n with respect to x is,

$$\boxed{\frac{\mathrm{d}\left[x^n\right]}{\mathrm{d}x} = nx^{n-1}} \qquad \text{Eqn. 1-23}$$

Hopefully, you also remember Eqn. 1-15, namely that the integral of x^n is $x^{(n+1)}/(n+1)$. To convert Eqn. 1-23 to an integral, multiply both sides by dx. (Don't be scared, we've already divided by dx so why not multiply by it.) Next, reverse the derivative process. Take the integral of each side. This shows that the integral of nx^{n-1} is, x^n plus a constant, we call C.

$$\int \mathrm{d}x^n = \int nx^{n-1}\mathrm{d}x = x^n + C$$

Generalizing, if we have a function, **big** $F(x)$ and we know that its derivative is simply **little** $f(x)$, then we can reverse the process by integrating **little** $f(x)$ to recover **big** $F(x)$. This idea asserts that **any function is the integral of its own derivative** or that the integral is the ***antiderivative***. It is called…

The Fundamental Theorem of Calculus

$$\boxed{\frac{\mathrm{d}\left[F(x)+C_1\right]}{dx} = f(x) \quad \leftrightarrow \quad \int f(x)\mathrm{d}x = F(x) + C} \qquad \text{Eqn. 3-1}$$

Way back in Chapter 1, I pointed out that

1: The integral is graphed as an area.
2: When no bounds are set for the domain of x the integral (or area) is indefinite.
3: An unknown constant, C must be added to all indefinite integrals because you can't calculate the area if you only know the height (the function) but don't know the width or the bounds.
4. Integrating and Differentiating are reverse processes.

Derivatives have lost all memory of any *added* constant, C_1 that was part of the original function. (You are not the only one that forgets.) For example, $y_1(x) = \mathbf{1} + x^n$ and $y_2(x) = \mathbf{5} + x^n$ are two functions that differ by 4 but have identical derivatives, namely, nx^{n-1} because the derivative of any constant (such as 1 or 5) is 0. So, when we reverse the process and take the integral, we know there may be some lost constant that we have to restore. But we can't retrieve C_1 if we don't specify the domain of the integral because it has been forgotten, so we just leave it as an anonymous, undetermined C.

You also need more data to evaluate integrals than derivatives. If I asked you to evaluate the derivative of $\frac{1}{3}x^3$ with respect to x you would get x^2 and then have to ask me, "At what value of x?" If $x = 5$, then the derivative is $x^2 = (5)^2 = 25$. If I asked you to find the integral of x^2 you would have to ask me, "What are the values of the domain, $a \le x \le b$?" If I said, "Start at $x = 2$ and end at $x = 5$", then $a = 2$ and $b = 5$ in Eqn. 1-18 and the definite integral or area is, is $\frac{1}{3}(b^3 - a^3) = \frac{1}{3}(5^3 - 2^3) = 117/3$.

When Eqn. 1-18 is rearranged and generalized to any continuous function, $f(x)$, it transforms to another form of the Fundamental Theorem of Calculus,

$$\boxed{\int_0^b f(x)\mathrm{d}x = \int_0^a f(x)\mathrm{d}x + \int_a^b f(x)\mathrm{d}x \quad \Leftrightarrow \quad F(b) - F(a) = \int_a^b f(x)dx} \qquad \text{Eqn. 3-2}$$

Eqn. 3-2 means that the area of a function over a total domain (from $x = 0$ to $x = b$, where $b > a > 0$) is the sum of the areas over the partial domains (from $x = 0$ to $x = a$ plus from $x = a$ to $x = b$), as Figure 3-1 shows. **Note: Limits are color coded to get your attention.**

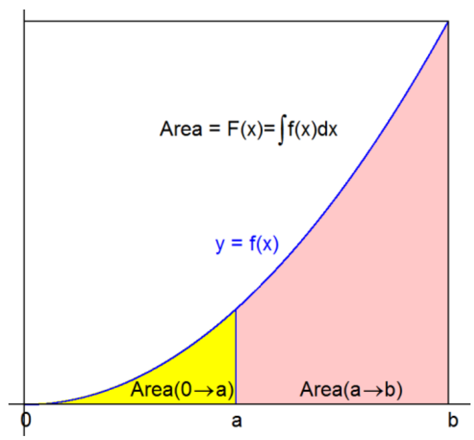


Figure 3-1 The total area (integral) is the sum of the areas (integrals) of each segment.

Summarizing and Reviewing

1: When no values of x are given above and below the integral sign, the integral always has an indefinite constant, C, and is called an indefinite integral. Thus, the indefinite integral of nx^{n-1} is $x^n + C$.

2: When upper and lower bounds of the independent variable, x are shown above and below the integral sign the integral is called a definite integral, and it can be evaluated because it can always be depicted by an area whose width is specified and whose height is given by the function to be integrated.

Problem: Evaluate the integral of $f(x) = 5x^4$ over the domain $1 \le x \le 3$.

Solution: The integral is $F(x) = x^5$. Then apply Eqn. 3-2. The definite integral becomes

$$\int_1^3 5x^4 dx = \int_0^3 5x^4 dx - \int_0^1 5x^4 dx = 3^4 - 1^4 = 243$$

Problem: Calculate the integral of x^2 for the domain, $2 \le x \le 5$.

Solution: There is a slight twist here. The derivative of x^3 with respect to x is $3x^2$.

Therefore, going backwards, the integral of $3x^2$ is x^3. To get the integral of x^2 we must divide by 3, which yields $⅓x^3$. Thus, the definite integral becomes

$$\int_2^5 x^2 \mathrm{d}x = \int_0^5 x^2 \mathrm{d}x - \int_0^2 x^2 \mathrm{d}x = \frac{5^3}{3} - \frac{2^3}{3} = \frac{98}{3}$$

Indefinite Integrals of Some Functions We Already Know

It is time to list indefinite integrals (antiderivatives) of functions we already know. We do this by applying the Fundamental Theorem of Calculus.

$$\frac{\mathrm{d}\left[x^n + C_1\right]}{\mathrm{d}x} = ne^{n-1} \quad \leftrightarrow \quad \int x^n \mathrm{d}x = \frac{x^{n+1}}{n+1} + C \qquad \text{Eqn. 3-3}$$

$$\frac{\mathrm{d}\left[\sin(x) + C_1\right]}{\mathrm{d}x} = \cos(x) \quad \leftrightarrow \quad \int \cos(x)\mathrm{d}x = \sin(x) + C \qquad \text{Eqn. 3-4}$$

$$\frac{\mathrm{d}\left[\cos(x) + C_1\right]}{\mathrm{d}x} = -\sin(x) \quad \leftrightarrow \quad \int \sin(x)\mathrm{d}x = -\cos(x) + C \qquad \text{Eqn. 3-5}$$

$$\frac{\mathrm{d}\left[\tan(x) + C_1\right]}{\mathrm{d}x} = \sec^2(x) \quad \leftrightarrow \quad \int \sec^2(x)\mathrm{d}x = \tan(x) + C \qquad \text{Eqn. 3-6}$$

$$\frac{\mathrm{d}\left[\ln(x) + C_1\right]}{\mathrm{d}x} = \frac{1}{x} \quad \leftrightarrow \quad \int \frac{\mathrm{d}x}{x} = \ln(x) + C \qquad \text{Eqn. 3-7}$$

$$\frac{\mathrm{d}\left[e^{kx} + C_1\right]}{\mathrm{d}x} = ke^{kx} \quad \leftrightarrow \quad k\int e^{kx} \mathrm{d}x = e^{kx} + C \qquad \text{Eqn. 3-8}$$

$$\frac{\mathrm{d}\left[\sin^{-1}(x) + C_1\right]}{\mathrm{d}x} = \frac{1}{\sqrt{1-x^2}} \quad \leftrightarrow \quad \int \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \sin^{-1}(x) + C \qquad \text{Eqn. 3-9}$$

$$\frac{\mathrm{d}\left[\tan^{-1}(x) + C_1\right]}{\mathrm{d}x} = \frac{1}{\sqrt{1+x^2}} \quad \leftrightarrow \quad \int \frac{\mathrm{d}x}{\sqrt{1+x^2}} = \tan^{-1}(x) + C \qquad \text{Eqn. 3-10}$$

Some of these integrals, especially Eqn. 3-9 and Eqn. 3-10 are real terrors to all Calculus students because almost the only way to figure them out is if you were lucky enough to <u>remember</u> what functions they were derivatives of. You will pay dearly if you forget them because the professor will almost certainly throw one of these integrals at you on your test.

3.2 Perfect Differentials and their Integrals

Any expression that has the form, d[f(x)], such as d[e^{kx}sin(x)] is called a perfect differential. As a counterexample, x·d[sin(x)] is not a Perfect Differential because x is not being differentiated. The Fundamental Theorem of Calculus guarantees that the integral of a perfect differential is the function itself (plus a constant).

Many expressions are disguised forms of perfect differentials. If we can transform an integrand (the function inside the integral) into a perfect differential we can integrate it immediately. Two examples are,

$$\boxed{2x \cdot \mathrm{d}x = \mathrm{d}\left[x^2\right] \qquad \& \qquad 2\cos(x) \cdot \mathrm{d}\left[\sin(x)\right] = \mathrm{d}\left[\sin^2(x)\right]}$$

Here is a trickier example.

Problem: Find the integral of tan(*x*) with respect to *x*.
Solution: At first glance it seems that we should know the integral of tan(*x*) immediately. But when we think for a moment or two, we realize that we don't. However, we do know integrals of sines and cosines, so hopefully the trick is to express the tangent as sin(x)/cos(x). We now convert tan(*x*)·d*x* to a perfect differential with the help of Eqn. 2-23.

$$\int \tan(x)\mathrm{d}x = \int \frac{\sin(x)\mathrm{d}x}{\cos(x)} = \int \frac{\mathrm{d}\left[-\cos(x)\right]}{\cos(x)} = -\int \mathrm{d}\left[\ln(\cos(x))\right]$$

The integrand on the right is a perfect differential, so we can integrate immediately.

$$\boxed{\int \tan(x)\mathrm{d}x = -\ln(\cos(x)) + C} \qquad \text{Eqn. 3-11}$$

3.3 Reversing Chain Gang Rules for Integrals of Complicated Functions

Finding integrals of complicated functions use the same rules (e. g., product, quotient, and chain rules) that we use to find derivatives of complicated functions. Once again, we combine these rules with the Fundamental Theorem of Calculus.

$$\mathrm{d}[uv] = u\mathrm{d}v + v\mathrm{d}u \quad \leftrightarrow \quad [uv] + C = \int \mathrm{d}[uv] = \int u\mathrm{d}v + \int v\mathrm{d}u \qquad \text{Eqn. 3-12}$$

$$\mathrm{d}\left[\frac{u}{v}\right] = \frac{v\mathrm{d}u - u\mathrm{d}v}{v^2} \quad \leftrightarrow \quad \left[\frac{u}{v}\right] + C = \int \mathrm{d}\left[\frac{u}{v}\right] = \int \frac{\mathrm{d}u}{v} - \int \frac{u\mathrm{d}v}{v^2} \qquad \text{Eqn. 3-13}$$

$$\mathrm{d}u\left[u(v(x))\right] = \frac{\mathrm{d}u}{\mathrm{d}v}\mathrm{d}v \quad \leftrightarrow \quad u(v(x)) + C = \int \mathrm{d}\left[u(v(x))\right] = \int \frac{\mathrm{d}u}{\mathrm{d}v}\mathrm{d}v \qquad \text{Eqn. 3-14}$$

WARNING: From now on, if I forget the constant, *C* in indefinite integrals, put it in yourself and don't even think of suing me! And please, please, please, don't make me warn you again!

Integration by Parts I: Indefinite Integrals of the Product Rule

Using Eqn. 3-12 (the Product Rule) to find integrals is called Integration by Parts. This is one of the most useful and powerful of all integration techniques and you can be guaranteed it will appear on a Calculus test. The Product Rule operates by transforming unfriendly versions of functions to friendlier versions.

For example, consider the integral, $\int$sin(*x*)·cos(*x*)d*x*. This may look difficult but it will simplify almost immediately. The job, and too often the trick, is to find *u*, *v*, d*u*, and d*v*. The integrand is clearly a product, and you might see that…

$$u = \sin(x) \qquad \mathrm{d}v = \cos(x)dx = \mathrm{d}[\sin(x)]$$
$$\mathrm{d}u = \cos(x)\mathrm{d}x \qquad v = \sin(x)$$

Now that we have identified *u*, *v*, d*u*, and d*v* we are ready to use the Product Rule

$$\int \underset{u}{\sin(x)}\underbrace{\cos(x)\mathrm{d}x}_{\mathrm{d}v} = \int \mathrm{d}\underbrace{\left[\sin^2(x)\right]}_{\mathrm{d}(uv)} - \int \underset{v}{\sin(x)}\underbrace{\cos(x)\mathrm{d}x}_{\mathrm{d}u}$$

Notice that the integral on the far right is exactly the same as the integral on the left. What luck! (It only happens to Calculus teachers.) Therefore, adding it to both sides yields,

$$2\int \sin(x)\cos(x)\mathrm{d}x = \int \mathrm{d}\left[\sin^2(x)\right]$$

But since the integral of a perfect differential is simply the function itself…

$$\int \mathrm{d}\left[\sin^2(x)\right] = \sin^2(x) + C$$

The final result is that

$$\boxed{\int \sin(x)\cos(x)\mathrm{d}x = \frac{1}{2}\sin^2(x) + C} \qquad \text{Eqn. 3-15}$$

Oops! We could have found this result sooner by simply using the chain rule if only we realized that cos(x)dx = d[sin(x)]. This means that the integrand of Eqn. 3-15 is a perfect differential because it has the form, $u\mathrm{d}u = \mathrm{d}[½u^2]$, where $u = \sin(x)$, so the integral is $½u^2 + C$.

The chain rule greatly increases our integrating power. Imagine we raise sin(x) to the n^{th} power and integrate $\sin^n(x)\cdot\cos(x)\mathrm{d}x$! The chain rule simplifies the integral to,

$$\int \sin^n(x)\cos(x)\mathrm{d}x = \int \underset{u^n}{\sin^n(x)}\underbrace{\mathrm{d}[\sin(x)]}_{\mathrm{d}u} = \underbrace{\int \mathrm{d}\left[\frac{\sin^{n+1}(x)}{n+1}\right]}_{= \, \mathrm{d}[u^{n+1}/(n+1)]}$$

Using the Fundamental Theorem of Calculus yields,

$$\boxed{\int \sin^n(x)\cos(x)\mathrm{d}x = \frac{\sin^{n+1}(x)}{n+1} + C} \qquad \text{Eqn. 3-16}$$

Similarly, you should be able to prove that

$$\boxed{\int \cos^n(x)\sin(x)\mathrm{d}x = -\frac{\cos^{n+1}(x)}{n+1} + C} \qquad \text{Eqn. 3-17}$$

Don't forget that minus sign!

A little more painful is the integral, $\int\cos^2(x)\mathrm{d}x$. This really requires genius, or at least ingenuity and luck. Luck or ingenuity consists in 1: splitting the $\cos^2(x)$ and choosing $u = \cos(x)$ and $\mathrm{d}v = \cos(x)\mathrm{d}x = \mathrm{d}[\sin(x)]$ in the integration by parts. Then, 2: we need to remember both Pythagoras and the multiple angle formula for $\sin(2x)$. The work follows.

$$\begin{aligned}\int \cos^2(x)\mathrm{d}x = \int \cos(x)\mathrm{d}[\sin(x)] &= \int \mathrm{d}[\cos(x)\sin(x)] - \int [-\sin^2(x)]\mathrm{d}x \\ &= \cos(x)\sin(x) + \int [1-\cos^2(x)]\mathrm{d}x\end{aligned}$$

Rearranging, setting $\cos(x)\cdot\sin(x) = \frac{1}{2}\sin(2x)$ and dividing by 2 yields the result,

$$\boxed{\int \cos^2(x)\mathrm{d}x = \frac{\sin(2x)}{4} + \frac{x}{2} + C} \qquad \text{Eqn. 3-18}$$

Some functions, such as $e^{kx}\cdot\cos(x)$, must be integrated by parts twice or more. This is not surprising considering that $\sin(x)$ and $\cos(x)$ reappear only after differentiating twice. For the first integral by parts, set $u = e^{kx}$ and $\mathrm{d}v = \cos(x)\mathrm{d}x = \mathrm{d}[\sin(x)]$. This yields,

$$\int e^{kx}\cos(x)\mathrm{d}x = \int \mathrm{d}\left[e^{kx}\sin(x)\right] - \boxed{\int ke^{kx}\sin(x)\mathrm{d}x}$$

The second integration by parts involves only the last (red shaded) integral on the right

$$\boxed{\int ke^{kx}\sin(x)\mathrm{d}x} = \int \mathrm{d}\left[ke^{kx}(-\cos(x))\right] - \int \left[k^2e^{kx}(-\cos(x))\right]\mathrm{d}x$$

Substituting this result and using the Fundamental Theorem of Calculus, yields,

$$\boxed{\int e^{kx}\cos(x)\mathrm{d}x = \frac{e^{kx}\sin(x) + ke^{kx}\cos(x)}{(k^2+1)} + C} \qquad \text{Eqn. 3-19}$$

We can also find the integral, $\int\ln(x)\mathrm{d}x$ using integration by parts. To do this, set $u = \ln(x)$ and $\mathrm{d}v = \mathrm{d}x$. Then, $\mathrm{d}u = \mathrm{d}x/x$ and $v = x$. Integrating by parts yields,

$$\boxed{\int \ln(x)\mathrm{d}x = \int \mathrm{d}[x\ln(x)] - \int x\frac{\mathrm{d}x}{x} = x\ln(x) - x + C} \qquad \text{Eqn. 3-20}$$

Once again, that turned out to be pretty simple. I wish I had thought of it! Of course, we will never know how much struggle and luck it took to discover this centuries old equation.

Integration by Parts II: Some Definite Integrals of the Product Rule

It is now time to give a few examples of using integration by parts to evaluate definite integrals when we have the bounds on the independent variable (usually x).

Problem: Evaluate the integral of $f(x) = \sin(x)\cdot\cos(x)$ for the domain, $0 \le x \le \pi/2$.
Solution: Begin by recalling Eqn. 3-15. We then write

$$\int_0^{\pi/2} \sin(x)\cos(x)\mathrm{d}x = \frac{1}{2}\int_0^{\pi/2} \mathrm{d}\left[\sin^2(x)\right] = \frac{\sin^2(\pi/2)}{2} - \frac{\sin^2(0)}{2} = \frac{1}{2}$$

What a simple result! Could you have anticipated it? I couldn't when I took Calculus!

Problem: Evaluate the integral of $f(x) = e^{x/2}\cos(x)$ for the domain, $0 \le x \le \pi$.
Solution: Begin by recalling Eqn. 3-19. Then, $k = ½$, and we write,

$$\int_0^{\pi} e^{\frac{x}{2}}\cos(x)\mathrm{d}x = \frac{1}{1.25}\left[\left\{e^{\frac{\pi}{2}}\sin(\pi) - e^0\sin(0)\right\} + 0.5\left\{e^{\frac{\pi}{2}}\cos(\pi) - e^0\cos(0)\right\}\right]$$
$$= 0.8\left[\{0-0\} + 0.5\left\{-e^{\pi/2} - 1\right\}\right] \approx -2.32$$

3.4 Negative and Positive Areas

You may find that the above result is unexpected and weird! How can area be negative? The answer is, easily. Simply think of your bank account or your parents' bank account, if they are paying your tuition, room and board.

Area is negative if the function is negative and the width is positive. In Figure 3-2 the function is positive between $x = 0$ and $x = \pi/2$, but then is more negative for $x > \pi/2$. The integral gives the net area above the x axis (the positive area minus the negative area) over the entire region.

Warning: The width is taken as positive if the integral's upper bound is greater than its lower bound. **Note**: The color coded bounds, a, and b, of both integrals are reversed in Eqn. 3-21.

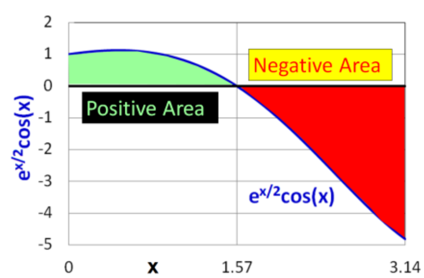


Figure 3-2 Definite integrals give net area above the x axis, i. e., green minus red areas.

$$\int_a^b f(x)\mathrm{d}x = F(b) - F(a) = -\left[F(a) - F(b)\right] = -\int_b^a f(x)\mathrm{d}x \qquad \text{Eqn. 3-21}$$

Thus, the definite integral reverses sign when the order of the bounds is reversed.

3.5 Integrating Obnoxious Fractions, I: Retrieving Logs

Memory is always paramount when trying to find integrals. When functions involve fractions, memory is far more important.

Fractions are always difficult to deal with, especially when constants are mixed with variables. Consider how easy it is to integrate $f(x)\mathrm{d}x$ when $f(x) = ax + b$. Simply integrate ax and b separately, add the two integrals, and then add that pesky unknown constant, C.

$$\int (ax+b)\mathrm{d}x = \int ax\mathrm{d}x + \int b\mathrm{d}x = \frac{ax^2}{2} + bx + C$$

Now, try to integrate f(x) = 1/(ax + b)! You must begin by suspecting that f(x) is similar to 1/x. Then, memory tells you that its integral is ln(x). Assuming that you are correct, all you have to do is mess with the constants in the fraction. There are at least two ways to skin this cat. In both cases we will do the definite integral, for reasons that will become apparent. The first is to separate the constant, a, from x and place it outside the integral.

$$\int_{x_1}^{x_2} \frac{\mathrm{d}x}{(ax+b)} = \frac{1}{a}\int_{x_1}^{x_2} \frac{\mathrm{d}x}{(x+b/a)} = \frac{1}{a}\ln\left(x_2 + \frac{b}{a}\right) - \frac{1}{a}\ln\left(x_1 + \frac{b}{a}\right) \qquad \text{Eqn. 3-22}$$

The second is to incorporate a into the differential d[ax]. This yields,

$$\int_{x_1}^{x_2} \frac{\mathrm{d}x}{(ax+b)} = \frac{1}{a}\int_{x_1}^{x_2} \frac{\mathrm{d}[ax]}{(ax+b)} = \frac{1}{a}\ln(ax_2 + b) - \frac{1}{a}\ln(ax_1 + b) \qquad \text{Eqn. 3-23}$$

Eqn. 3-22 and Eqn. 3-23 may look different but despite appearances (which are deceiving - certainly for logarithms) they are identical! Here is the proof! Recall that ln(ax) = ln(x) + ln(a). Thus, [ln(ax_2 + b)] – [ln(ax_1 + b)] = [ln(x_2 + b/a) + ln(a)] – [ln(x_1 + b/a) + ln(a)], and the two integrals are identical.

3.6 Integrating Obnoxious Fractions, II: Making Perfect Differentials

Which of the two following two integrals do you think are easier to evaluate?

$$\int \frac{\mathrm{d}x}{(x^2+a^2)} \quad or \quad \int \frac{x\mathrm{d}x}{(x^2+a^2)} \quad ?$$

These may look alike, but look carefully. The second integrand has an extra factor of x in the numerator. That factor actually makes the integral easier because the variable in the denominator is x^2 and the numerator, $x\mathrm{d}x$ = ½d[x^2] = ½d[x^2 + a^2]. This means the integrand can be transformed into a perfect differential. To do this, set

$$u = x^2 + a^2 \quad \Rightarrow \quad du = 2xdx$$

With this substitution, we manipulate the integrand into the form of an old friend, the perfect differential of the natural log. Then we can integrate immediately.

$$\int \frac{x\mathrm{d}x}{(x^2+a^2)} = \frac{1}{2}\int \frac{\mathrm{d}u}{u} = \frac{1}{2}\int \mathrm{d}[\ln(u)] = \frac{1}{2}\ln(x^2+a^2) \qquad \text{Eqn. 3-24}$$

The integral dx/(x^2 + a^2) seems easier until we notice it is not a perfect differential. But if you strain your memory and imagination the integrand might remind you of the derivative of the inverse tangent, except for the constant a^2 in place of 1. That shouldn't be too big a problem because we can divide numerator and denominator by a^2 and then set u = x/a. The integral then becomes a multiple of the inverse tangent

$$\int \frac{\mathrm{d}x}{(x^2+a^2)} = \frac{1}{a}\int \frac{1}{\left(\dfrac{x^2}{a^2}+1\right)}\mathrm{d}\left[\frac{x}{a}\right] = \frac{1}{a}\int \frac{\mathrm{d}u}{(u^2+1)} = \frac{1}{a}\int \mathrm{d}\left[\tan^{-1}(u)\right]$$

So, finally, the integral of dx/(x^2 + a^2) is,

$$\boxed{\int \frac{dx}{(x^2+a^2)} = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right) + C} \qquad \text{Eqn. 3-25}$$

3.7 Integrating Obnoxious Fractions, III: Substituting Trigonometric Functions

Algebraic expressions that involve 1/(x^2 + 1) and 1/(1 - x^2)^{½} can be integrated if x is replaced by tan(u) and sin(u) respectively. There are two motivations for making these substitutions. First, and most important, your professor will almost certainly test you on them. Second, remember that tan^2(u) + 1 = sec^2(u) and (1 - sin^2(u))^{½} = cos(u). These identities not only simplify the function, they simplify the integration.

Your professor can get even trickier by asking you to integrate algebraic expressions that involve 1/(x^2 + a^2). Since we just finished integrating dx/(x^2 + a^2), you know the trick. First set u = x/a. That transforms (x^2 + a^2) to a^2(u^2 + 1). Then set u = tan (v) and go to town. Here is a list of four of these combinations and their integrals.

$$\boxed{\frac{1}{(x^2+a^2)} \Rightarrow \frac{x}{a} = \tan(v) \mapsto \int \frac{\mathrm{d}x}{(x^2+a^2)} = \frac{1}{a}\int \frac{\mathrm{d}[\tan(v)]}{(\tan^2(v)+1)} = \frac{v}{a} = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right)}$$

$$\boxed{\frac{1}{(x^2+1)} \Rightarrow x = \tan(v) \mapsto \int \frac{\mathrm{d}x}{(x^2+1)} = \int \frac{\mathrm{d}[\tan(v)]}{(\tan^2(v)+1)} = \int \frac{\sec^2(v)\mathrm{d}v}{\sec^2(v)} = v = \tan^{-1}(x)}$$

$$\boxed{\frac{1}{\sqrt{1-x^2}} \Rightarrow x = \sin(v) \mapsto \int \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \int \frac{\mathrm{d}[\sin(v)]}{\cos(v)} = \int \frac{\cos(v)\mathrm{d}v}{\cos(v)} = v = \sin^{-1}(x)}$$

$$\boxed{\frac{1}{\sqrt{a^2-x^2}} \Rightarrow \frac{x}{a} = \cos(v) \mapsto \int \frac{\mathrm{d}x}{\sqrt{a^2-x^2}} = \int \frac{\mathrm{d}[\cos(v)]}{\sin(v)} = -v = -\cos^{-1}\left(\frac{x}{a}\right)}$$

Now, let's go to town! Find the integral,

$$\int \frac{dx}{(x^2+a^2)^2}$$

1: Replace *x*/*a*, i. e., set *x*/*a* = u

$$\int \frac{\mathrm{d}x}{(x^2+a^2)^2} = \frac{1}{a^3}\int\left[\left(\frac{x}{a}\right)^2+1\right]^{-2}\mathrm{d}\left[\frac{x}{a}\right] = \frac{1}{a^3}\int\frac{\mathrm{d}u}{\left(u^2+1\right)^2}$$

2: Replace *u*, i. e., set *u* = tan(*v*).

$$\frac{1}{a^3}\int\frac{\mathrm{d}u}{\left(u^2+1\right)^2} = \frac{1}{a^3}\int\frac{\mathrm{d}[\tan(v)]}{\left(\tan^2(v)+1\right)^2} = \frac{1}{a^3}\int\frac{\sec^2(v)\mathrm{d}v}{\sec^4(v)} = \frac{1}{a^3}\int\cos^2(v)\mathrm{d}v$$

Great! $\int\cos^2(v)\mathrm{d}v$ is identical to Eqn. 3-18, with *v* in place of *x*.

$$\boxed{\frac{1}{a^3}\left[\int\cos^2(v)\mathrm{d}v = \frac{\sin(2v)}{4}+\frac{v}{2}\right]}$$

3. Convert *v* back to *u*.

Warning: I will take a step that leads to a dead end and then have to backtrack one step! There is nothing wrong with taking wrong steps that seem natural and then backtracking. Remember the mazes you did when you were a kid? After a wrong step you simply backtracked to find the right path.

$$\frac{1}{a^3}\left[\int\cos^2(v)\mathrm{d}v = \boxed{\frac{\sin(2v)}{4}}+\frac{v}{2} = \boxed{\frac{\sin\left[2\tan^{-1}(u)\right]}{4}}+\frac{\tan^{-1}(u)}{2}\right]$$

The term in the green box is OK but if we convert *v* directly to *u*, this term converts to the term in the pink box. This is a dead end as far as I can see because I have no idea of what sin[2·tan^{-1}(*u*)], i. e., the sine of the inverse tangent means. Do you?

So, take one step back to the term in the green box. The aim is to express sin(2*v*) in terms of tan(*v*) because it converts directly back to *x* via the equations, tan(*v*) = *u* = *x*/*a*. The flash of genius to do this is to use the multiple angle formula for sin(2*v*). Then it still takes the following two lines of trigonometry!

$$\frac{1}{4}\sin(2v) = \frac{1}{2}\sin(v)\cos(v) = \frac{1}{2}\frac{\sin(v)}{\cos(v)}\cos^2(v) = \frac{1}{2}\tan(v)\frac{1}{\sec^2(v)} =$$

$$= \frac{1}{2}\frac{\tan(v)}{\tan^2(v)+1} = \frac{1}{2}\frac{u}{u^2+1} = \frac{1}{2}\left(\frac{x}{a}\right)\left[\left(\frac{x}{a}\right)^2+1\right]^{-1}+C$$

That was easy, wasn't it? All we needed was that flash.

4: To finish the integral, convert *u* back to *x*.

Combining everything, we find...

$$\boxed{\int \frac{\mathrm{d}x}{(x^2+a^2)^2} = \frac{1}{2}\left(\frac{x}{a}\right)\left[\left(\frac{x}{a}\right)^2+1\right]^{-1} + \frac{1}{2}\tan^{-1}\left(\frac{x}{a}\right)+C} \qquad \text{Eqn. 3-26}$$

Voila!! This is almost as tough as it gets. If you have followed this derivation with my false step, then you are approaching genius status and I praise you highly. After all, when I review what I have written I have trouble following it myself...unless I read very slowly.

3.8 Integrating Obnoxious Fractions, IV: Partial Fractions

What in the world is a partial fraction? I distinctly remember that integrating partial fractions was one of the topics in Calculus that I hated the most as a student. It might have been easier if it had been presented in reverse order, which is what I am going to do now.

Imagine that your professor asked you on a test to integrate the following fractions.

$$f_1(x) = \frac{2}{(x+3)} + \frac{3}{(x-1)}$$

By this point we have seen so many logs that it should be a piece of cake! Integrate both fractions separately and each integral is a natural log

$$2\int \frac{\mathrm{d}x}{(x+3)} + 3\int \frac{\mathrm{d}x}{(x-1)} = 2\ln(x+3) + 3\ln(x-1)$$

But, what if I asked you to integrate the single fraction below?

$$f_2(x) = \frac{5x+7}{(x^2+2x-3)}$$

Huh! This seems much more difficult until I tell you that $f_1(x) = f_2(x)$ so their integrals are also equal. Clearly, the technique for integrating the complicated fraction, $f_2(x)$ is to break it down into the two simpler fractions that make up $f_1(x)$. When a complicated fraction can be broken down into simpler fractions, the simpler fractions are called partial fractions.

In my High School algebra class, we combined simpler, partial fractions into more complex fractions (without ever calling the originals partial fractions), but I don't recall ever trying to separate complex fractions into the partial fractions. Then in Calculus, they started right off finding the partial fractions of more complex fractions without connecting what they were doing with what we had done in High School. It seems that they did everything they could to make the process more difficult and mysterious. No wonder I hated it.

So, let's briefly review how to combine the partial fractions that make up $f_1(x)$ into the fraction that appears as $f_2(x)$. Start by cross multiplying to get a common denominator.

$$f_1(x) = \frac{2}{(x+3)} + \frac{3}{(x-1)} = \frac{2}{(x+3)}\frac{(x-1)}{(x-1)} + \frac{3}{(x-1)}\frac{(x+3)}{(x+3)} = \frac{5x+7}{(x^2+2x-3)} = f_2(x)$$

Now reverse the process and learn how to start with a more complicated fraction such as

$f_2(x)$ and break it down into its simpler partial fractions. This is a tougher job.

Step #1: Factor the denominator of the complicated function. Complicated fractions can only be broken down into partial fractions if we can factor the denominator of the complicated fraction! (No one ever told me that!).

$$\left(x^2+2x-3\right)=(x+3)(x-1)$$

Step #2 and #3: Assign unknown constants (call them *A* and *B*) to the numerators of the partial fractions. Then cross multiply to produce a common denominator.

$$f_2(x)=\frac{5x+7}{\left(x^2+2x-3\right)}=\frac{A}{(x+3)}+\frac{B}{(x-1)}=\frac{A}{(x+3)}\frac{(x-1)}{(x-1)}+\frac{B}{(x-1)}\frac{(x+3)}{(x+3)}$$

Step #4: Combine the two fractions on the right

$$\frac{A}{(x+3)}\frac{(x-1)}{(x-1)}+\frac{B}{(x-1)}\frac{(x+3)}{(x+3)}=\frac{(A+B)x+3B-A}{\left(x^2+2x-3\right)}=\frac{5x+7}{\left(x^2+2x-3\right)}$$

Step #5: At this point we need only set the numerators equal.

$$(A+B)x+3B-A=5x+7$$

This requires a final flash of genius (highlighted by color coding)! Since this equation must be true for all values of the variable, *x*, it is really two equations for the two unknowns, *A* and *B*, namely,

$$[A+\ B=5]x \qquad\qquad -A+3B=7$$

Adding these two equations yields $4B = 12$ so $B = 3$. Substituting for *B* in the first equation then yields $A + 3 = 5$ or $A = 2$. Thus the problem is solved.

Trickier Partial Fractions

Think that you are done? Not quite. Naturally, there are at least two trickier problems involving partial fractions. They occur when one of the factors of the denominator is a quadratic such as $(x^2\text{-}1)$ or is repeated, such as $(x\text{-}1)^2$ (hence also quadratic). Then the rule is...The numerator of any partial fraction that is quadratic has the form $Ax + B$.

Problem: Prove that the full fraction on the left hand side equals the partial fractions on the right hand side

$$\frac{7x^2-5x+8}{\left(x^3-x^2+x-1\right)}=\frac{2x-3}{\left(x^2+1\right)}+\frac{5}{(x-1)}$$

Solution: Imagine that we did not know the numerators of the partial fractions, but we do know their denominators because we factored the big fraction.

$$\frac{7x^2-5x+8}{\left(x^3-x^2+x-1\right)}=\frac{Ax+B}{\left(x^2+1\right)}+\frac{C}{(x-1)}$$

We must solve for the 3 unknowns, *A*, *B*, and *C*. Once again begin to solve by factoring (which we have already done above) and then cross multiplying. This yields,

$$\frac{(Ax+B)}{(x^2+1)}\frac{(x-1)}{(x-1)}+\frac{C}{(x-1)}\frac{(x^2+1)}{(x^2+1)}=\frac{(C+A)x^2+(B-A)x+(C-B)}{(x^3-x^2+x-1)}$$

Once again we equate numerators but now there are 3 separate equations, (one for the x^2 terms, one for the x terms and one for the constants alone) namely,

$$\begin{aligned} A \quad + \quad C &= \;\; 7 \\ -A+B \qquad &= -5 \\ -B+\; C &= \;\; 8 \end{aligned}$$

I won't plague you with the algebra here – you must plague yourself. One technique is to add all 3 equations, which by sheer luck eliminates both *A* and *B* and yields 2*C* = 10. Then, substitute for *C* in each of the 1st and 3rd equations to find that *A* = 2 and *B* = -3.

Finally, consider the second obnoxious case (in the center, with the answer on the left hand side) of finding partial fractions, namely when one of the factors of the denominator is repeated. The critical point is that a repeated linear factor is quadratic.

$$\frac{3x-1}{(x+2)^2}-\frac{2}{(x+1)}=\frac{x^2-6x-9}{(x^3-x^2+x-1)}=\frac{Ax+B}{(x+2)^2}+\frac{C}{(x+1)}$$

I'll let you do the rest because it is essentially a repeat of the example just above.

3.9 An Integral that Needs All the Tricks: (This One is Long, Long, Long!)

Here is an integral that seems easy enough until you start trying to do it.

$$F(x)=\int \left(a^2+x^2\right)^{0.5}\mathrm{d}x$$

I encountered this integral while writing Section 6.9. It is the length of the arc of a parabola. Then I looked back here to Chapter 3 to see if I had done it. But I hadn't. When I started trying to derive it, I quickly found it is a real bear that requires almost every technique (i. e., trick) in the book. As you will see very, very soon, it is so long and intorted that if your professor ever gives you this integral on a test, then you know that he or she is a either sadist who should be fired or a sage who should be promoted. In either case, you had better be prepared. Let's do it step by step and count steps just to prove how long it is.

First, however, to cut the suspense and keep all of us sane, I will give the solution to the integral, Eqn. 3-31.

$$\boxed{\int \left(a^2+x^2\right)^{\frac{1}{2}}\mathrm{d}x=\frac{1}{2}\left[\sqrt{x^2+a^2}\right]x+\frac{a^2}{2}\left[\ln\sqrt{x^2+a^2}+x\right]} \qquad \text{Eqn. 3-31}$$

That doesn't look *toooo* complicated. And it is not the first time that the natural log has shown up seemingly out of nowhere - recall Eqn. 3-11. To check that Eqn. 3-31 is correct, I urge you to take the time to work backwards from the solution. Take the derivative of the right hand side and after a few minutes you will be able to show that it equals $(a^2 + x^2)^{0.5}$.

But if you didn't have the answer, here are the 10 steps you will need to find it. If you can find a shorter way, let me (and every other mathematician) know!

Step #1: Get the constant out of the integral by setting

$$x = au \quad \Rightarrow \quad x^2 = a^2u^2$$
$$\Rightarrow \quad \mathrm{d}x = a\mathrm{d}u$$

The integral then becomes,

$$\int \left(a^2 + x^2\right)^{\frac{1}{2}} \mathrm{d}x = a^2 \int \left(1 + u^2\right)^{\frac{1}{2}} \mathrm{d}u \qquad \text{Eqn. 3-27}$$

Step #2: I hope you remember that $(1 + u^2)$ suggests you set

$$u = \tan(y) \quad \Rightarrow \quad 1 + u^2 = \sec^2(y)$$
$$\Rightarrow \quad \mathrm{d}u = \mathrm{d}\left[\tan(y)\right] = \sec^2(y)\mathrm{d}y$$

Eqn. 3-27 then becomes,

$$\int \left(a^2 + x^2\right)^{\frac{1}{2}} \mathrm{d}x = a^2 \int \sec^3(y)\mathrm{d}y = a^2 \int \sec(y)\mathrm{d}\left[\tan(y)\right] \qquad \text{Eqn. 3-28}$$

Step #3: Integrate the right hand side of Eqn. 3-28 by parts.

$$a^2 \left\{ \int \sec^3(v)\mathrm{d}v = \int \sec(v)\mathrm{d}\left[\tan(v)\right] = \sec(v)\tan(v) - \int \tan(v)\mathrm{d}\left[\sec(v)\right] \right\}$$

Step #4: When we set $\mathrm{d}[\sec(y)] = \sec(y)\cdot\tan(y)\mathrm{d}y$, the last integral on the right hand side is gradually converted to two integrals

$$-\int \tan(v)\mathrm{d}\left[\sec(v)\right] = -\int \tan^2(v)\sec(v)\mathrm{d}v =$$

$$= -\int \left[\sec^2(v) - 1\right]\sec(v)\mathrm{d}v = \int \sec(v)\mathrm{d}v - \int \sec^3(v)\mathrm{d}v$$

Steps #5, and #6: Substitute into Eqn. 3-28, and rearrange. This yields the semifinal form

$$a^2 \int \sec^3(v)\mathrm{d}v = \frac{a^2}{2}\sec(v)\tan(v) + \frac{a^2}{2}\int \sec(v)\mathrm{d}v \qquad \text{Eqn. 3-29}$$

This integral doesn't give us a single break. We must now integrate $\sec(y)\mathrm{d}y$. I have seen two ways to do this (because I gave up trying myself). One involves an incredible trick while the other involves a smaller trick and partial fractions.

Step #7: The Incredible Trick
Simultaneously multiply and divide the integrand by [sec(y) + tan(y)]. It works but who would have thought of it????? Not moi!!!!!

$$\sec(y)\mathrm{d}y = \sec(y)\frac{\sec(y)+\tan(y)}{\sec(y)+\tan(y)}\mathrm{d}y = \frac{\sec^2(y)+\sec(y)\tan(y)}{\tan(y)+\sec(y)}\mathrm{d}y$$

Step #8: Observe that each term in the numerator of the right hand side is the derivative of the term directly below it in the denominator. This converts the integrand into the perfect differential,

$$\sec(y)\mathrm{d}y = \mathrm{d}\left[\ln\left\{\sec(y)+\tan(y)\right\}\right]$$

Thus, the integral of the secant is,

$$\int \sec(v)\mathrm{d}v = \ln\left[\tan(v)+\sec(v)\right]+C \qquad \text{Eqn. 3-30}$$

Steps #9 and #10: Clean up the mess and convert back to the original variable, x. Thus, aside from the usual constant

$$\boxed{\int \left(a^2+x^2\right)^{\frac{1}{2}}\mathrm{d}x = \frac{1}{2}\left[\sqrt{x^2+a^2}\right]x + \frac{a^2}{2}\left[\ln\sqrt{x^2+a^2}+x\right]+C} \qquad \text{Eqn. 3-31}$$

As I mentioned, we will use Eqn. 3-31 in Chapter 6 to find the arc length of a parabola.

An Alternate Technique
The other way to evaluate the integral of Eqn. 3-30 is to rewrite sec(y) as 1/cos(y). Then comes the major (but not stupendous) trick, which is to multiply both top and bottom by cos(y) and start playing the trigonometric substitution games. This yields,

$$\int \sec(v)\mathrm{d}v = \int \frac{1}{\cos(v)}\mathrm{d}v = \int \frac{\cos(v)}{\cos^2(v)}\mathrm{d}v = -\int \frac{\mathrm{d}\left[\sin(v)\right]}{1-\sin^2(v)}$$

With this last integral I hope you are beginning to feel as if you are on familiar territory. The natural thing to do next is to set sin(y) = z. Why is that natural? (Actually, nothing is natural about Calculus.) It is natural because then we can factor the denominator as, $1 - z^2 = (1 - z)(1 + z)$ so we can solve the integral by partial fractions.

$$-\int \frac{\mathrm{d}\left[\sin(y)\right]}{1-\sin^2(y)} = -\int \frac{\mathrm{d}z}{1-z^2} = -\frac{1}{2}\left[\int \frac{\mathrm{d}z}{1+z} + \int \frac{\mathrm{d}z}{1-z}\right] = -\frac{1}{2}\ln\left(\frac{1+z}{1-z}\right)+C$$

When you substitute backwards from z to y to u to x, and make some extra manipulations (and no careless errors), take my word for it that it will be the rightmost term of Eqn. 3-31.

3.10 Integrals, Averages, and Weighted Means: Center of Gravity
Just as noses were perfectly designed to fit glasses, integrals were perfectly designed to calculate averages of continuous functions. For the parabola, $y = x^2$, over the domain,

$0 \le x \le 1$, the integral equals ⅓, and that is also the arithmetic mean or average height.

We all know how to calculate means or arithmetic averages. Simply add all the numbers and divide by the number of terms. For example, my annual income is $100,000, yours is $1,000, an average plumber's is $499,000 and the average salary of a CEO who has driven a large public corporation into bankruptcy through eloquent incompetence and greed is $19,400,000 because he\she is a “job creator” (i. e., someone who fires many workers before floating away on golden parachutes). The total salary for the 4 people is 20 million. So, the mean salary is $5,000,000, which is not representative of anyone because the mean is distorted by one outsized member. This distortion is one reason that the median, or value of the 50^{th} percentile is often used to represent a typical number.

Weighted means are more involved. A standard example is test average in a class of 10 people, where 8 people (who happened to sit together) got 50% and the other 2 people (who sat together on the other side of the room) got 100%. (Did anyone cheat?) Class average is surely not 75% because more people got 50% than 100%. Each grade must be weighted or multiplied by the number of people earning that grade. The weighted mean is,

$$\overline{G} = \frac{8 \times 50\% + 2 \times 100\%}{8+2} = 60\%$$

The horizontal line or bar over $f_{wt}(x)$ is called an overbar and is the standard way to indicate a mean or a weighted mean.

One classical example of a weighted mean in physics is the Center of Mass. On Earth it is the balancing point and is often called the Center of Gravity. It is the mean of a distance multiplied or weighted by mass or weight. The balancing point does not occur in the geographical center of an object whose shape or density is asymmetric.

In general, a weighted mean is the mean of a function, f(x) that is multiplied by some weighting factor, w(x), and is defined by Eqn. 3-32.

$$\boxed{\overline{f_{wt}(x)} \equiv \frac{\sum w_i \cdot f_i}{\sum w_i} = \frac{\int w(x) \cdot f(x)\mathrm{d}x}{\int w(x)\mathrm{d}x}} \qquad \text{Eqn. 3-32}$$

Problem: Find the x value of the center of gravity of a triangular wedge, $y = x$ for $0 \le x \le 1$.
Solution: Distance f(x) = x is the function and height of the wedge, y = $w(x) = x$ is the weighting factor. Then Eqn. 3-32 becomes,

$$\overline{x_{cg}(y=x)} = \frac{\int_0^1 x^2 \mathrm{d}x}{\int_0^1 x\mathrm{d}x} = \frac{\frac{1^3}{3}}{\frac{1^2}{2}} = \frac{2}{3}$$

Problem: Find the x value of the center of gravity of the wedge, $y = e^{-4x}$ for $0 \le x \le 1$.
Solution: Use Eqn. 3-32 with $y = w(x) = e^{-4x}$. Hint: Integrate the numerator by parts.

$$\overline{x_{cg}(y=e^{-4x})} = \frac{\int_0^1 xe^{-4x}\mathrm{d}x}{\int_0^1 e^{-4x}\mathrm{d}x} = \frac{-\frac{1}{4}\left[e^{-4} - \left(-\frac{1}{4}\left(e^{-4}-1\right)\right)\right]}{\frac{\left(1-e^{-4}\right)}{4}} \approx \frac{0.0568}{0.245}$$

Thus $(e^{-4x})_{avg} \approx 0.245$ and $x_{cg}(e^{-4x}) \approx 0.0568/0.245 \approx 0.231$.

Figure 3-3 shows the x values of the centers of mass (indicated by the little triangles) of 3 objects. The center graph is the easiest. The weighting factor is constant so the balancing point is the midpoint, $x = 0.5$. The left graph has the same mass but clearly is heavier on its right side, so the balancing point is right of center at $x = 0.667$. The right graph decays exponentially at a rapid rate over the interval, $0 \le x \le 1$. The average value of the height is therefore less than 0.5. The balancing point is located relatively close to the left hand side (at $x \approx 0.231$) where the weighting factor is greatest (or tallest).

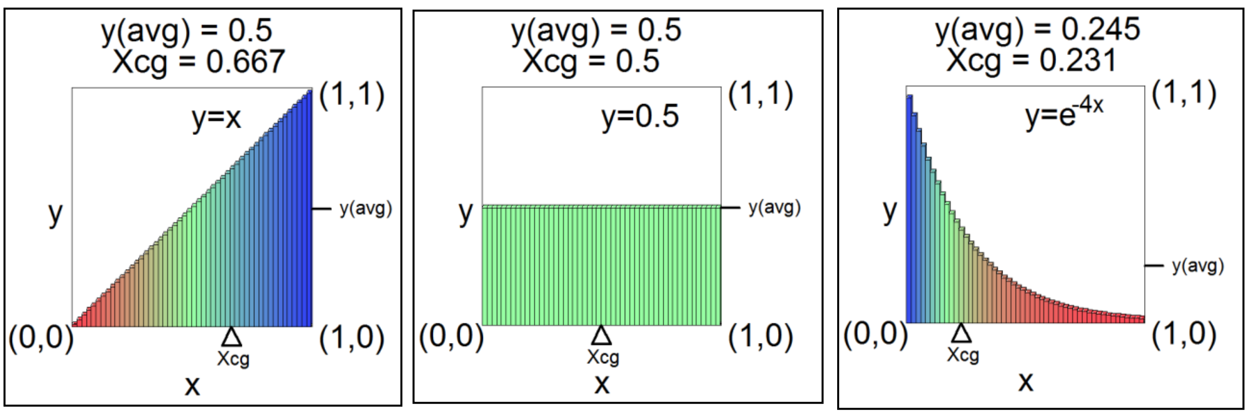


Figure 3-3 Means and centers of mass of 3 shapes (weighting factors) described in the text.

It is critical to find the center of gravity of any spinning component of an engine because that is the only point that will not cause the entire machine to wobble and possibly fail. You know that a washing machine wobbles and may even thump loudly in its spin cycle if the wet laundry is concentrated on one side of the basin.

3.11 Double Integrals: A New Way to Find Areas

So far, we have treated areas as the integrals or sums of a bunch of tall, skinny rectangles with width, Δx and height $y(x)$. We can also treat areas as the sum of a grid of tiny squares with width, Δx and height, $\Delta y = \Delta x$. This is shown in Figure 3-4 for the parabola, $y = x^2$ in the domain, $0 \le x \le 1$.

When we add squares we take two sums; first, the squares in each column and then all the columns. Thus one sum is nested inside the other. For the parabola, first fix x, then add the tiny squares in each column that extend from $y = 0$ to $y = x^2$. Finally, add all the columns from $x = 0$ to $x = 1$.

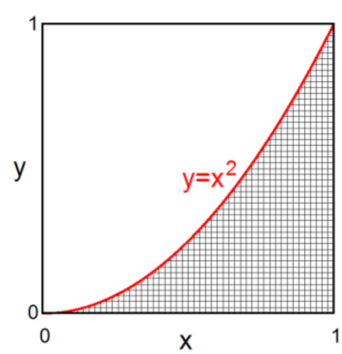


Figure 3-4 Find area by summing squares.

When $\Delta y \to 0$ and $\Delta x \to 0$, the two nested sums become two nested integrals (one with respect to *x* and one with respect to *y*), called a double integral. Just as the two sums must be calculated one at a time, so too, each integral of a double integral must be calculated one at a time. The order of integration does not matter; what is essential (i. e., tricky) is to get the correct limits of integration. The double integral can be written as,

$$\boxed{A = \int_{x_1}^{x_2}\left[\int_{y_1}^{y_2} \mathrm{d}y\right]\mathrm{d}x} \qquad \text{Eqn. 3-33}$$

I tried integrating both with respect to *x* first and with respect to *y* first. For the parabola of Figure 3-4 and most functions where $y = \mathrm{f}(x)$, I find it easier to integrate with respect to *y* first. In any case, the inner integral is always done first. At each value of *x*, the variable *y* ranges from 0 to x^2. Then *x* extends simply from 0 to 1. In that case, Eqn. 3-33 becomes,

$$A = \int_0^1\left[\int_0^{x^2} \mathrm{d}y\right]\mathrm{d}x = \int_0^1 x^2\mathrm{d}x = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3} \qquad \text{Eqn. 3-34}$$

Let's reverse the order and integrate with respect to *x* first, just to prove it can be done. Remember, that the tricky part is to get the correct limits of integration. Integrating with respect to *x* means that we count squares in each **row** holding *y* fixed, so that *x* ranges from $y^{1/2}$ to 1. (Note that on any row for $y > 0$ the squares start to the right of the parabola and not at $x = 0$.) The integral with respect to *y* ranges from $y = 0$ to $y = 1$ because there are rows at all values of *y* from 0 to 1. Because the order of integration has been reversed, the double integral looks and works out differently but arrives at the same result, $A = ⅓$.

$$A = \int_0^1\left[\int_{y^{0.5}}^1 \mathrm{d}x\right]\mathrm{d}y = \int_0^1 \left(1 - y^{0.5}\right)\mathrm{d}x = \left(y - \frac{y^{1.5}}{1.5}\right)\Bigg\|_0^1 = \frac{1}{3}$$

Double Integrals and Volumes

When we move from 2 to 3 dimensions, we move from areas to volumes. Then, instead of counting squares, we can weigh square columns of height, $z(x, y)$ as in Figure 3-5 to find volume. And, *z* can represent any quantity including density, mass, salary, age, etc.

We can calculate volume with these square columns using a 3-dimensional analog to Riemann Sums. In the limit as the width of the squares, $\Delta x = \Delta y \to 0$, this becomes a double integral for volume, *V*, given by Eqn. 3-35,

$$\boxed{V = \int_{x_1}^{x_2}\left[\int_{y_1}^{y_2} z(x,y)\mathrm{d}y\right]\mathrm{d}x} \qquad \text{Eqn. 3-35}$$

In Figure 3-5, I designed the height, $z(x, y)$ to be,

$$z = \begin{matrix} 0.3(1-x) + y^2 & y \le x^2 \\ 0 & y > x^2 \end{matrix}$$

so that it fits into a unit cube (length = width = height = 1, and V = 1) that extends from (x, y, z) = (0, 0, 0) to (1, 1, 1). Eqn. 3-35 has at least three other marvelous features. You can, 1: clearly see its shape, 2: integrate it to find V and, 3: compare V to that of the unit cube that confines it. The colored volume, V, fills only a modest fraction of the unit cube.

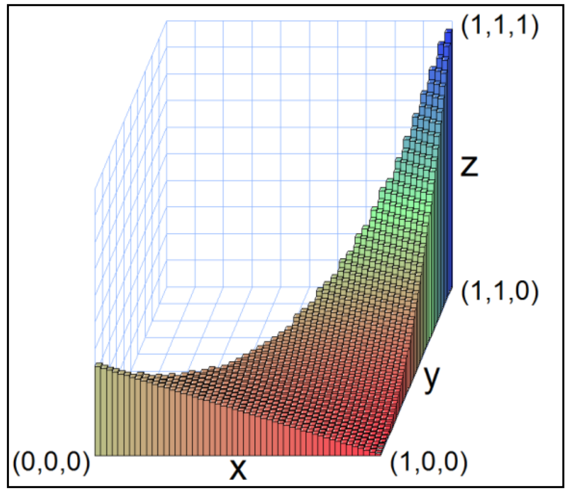


Figure 3-5 Volume of square columns bounded by $y = x^2$ and $z = 0.3(1-x)+y^2$.

Now, we calculate that V ≈ 0.131. Are you ready? It will not be so terrible. We keep the parabola, $y = x^2$ for the floor plan of the volume but now replace every square with a column of height, $z = 0.3(1 - x) + y^2$. The double integral is really easy if we do it step by step. Take the integral with respect to y first, holding x constant. Then integrate with respect to x. Because the volume has the same floor plan, the double integral has the same limits as in Eqn. 3-34, namely, $0 \le y \le x^2$ and $0 \le x \le 1$.

$$V = \int_0^1 \left[\int_0^{x^2} \left[0.3(1-x) + y^2 \right] dy \right] dx = \int_0^1 \left[0.3x^2(1-x) + \frac{x^6}{3} \right] dx = 0.3\left(\frac{1^3}{3} - \frac{1^4}{4} \right) + \frac{1^7}{3 \cdot 7} \approx 0.131$$

A Triple Integral Treat to Come

We have used both single and double integrals to find areas, which are 2-dimensional. We just used double integrals built of columns to find volumes, which are 3-dimensional. It makes sense that we can also use triple integrals (built of cubes or volume elements) to find volumes. That I postpone until Chapter 6 as a treat you are barely willing to wait for.

3.12 Numerical Integration: How to Integrate When You Can't

On my final exam in Fluid Dynamics I obtained an integral I couldn't solve. When I saw the Professor after the exam, he said, "Of course you couldn't do it. No one can. It is an Elliptic Integral. You should have approximated." Of course, he told us to solve and didn't tell us to approximate. I won't tell you and didn't tell him where he should have gone.

It was a lesson, nonetheless. When you get an integral that is too hard to do, or if you simply want to find the value of any integral - even an easy one - you can always approximate it real closely by integrating numerically. The technique, which I hope you remember from Section 1.2, is the Riemann Sum. (Remember, this technique is also called Quadrature.)

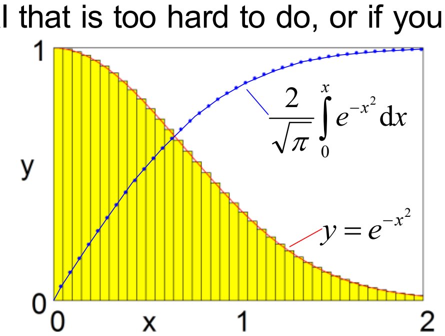


Figure 3-6 The Gaussian and its integral, the Error Function.

Here are some functions (integrands) that look innocent enough, but no exact, analytic functions exist for their integrals. Don't worry though, because every one of the integrals can be graphed and approximated.

The first is the integral of the Gaussian function, e^{-kx^2}, shown in Figure 3-6 (with k = 1) which is fundamental in statistics (see the Normal Distribution in Section 6.14). The name for the integral is the error function, erf(x). With a name like that it is no wonder it is important in statistics!

$$\boxed{\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-x^2}\,dx} \qquad \text{Eqn. 3-36}$$

The next function we can only integrate numerically is, $\sqrt{1-\varepsilon^2\sin^2(\theta)}$. This is the element of arc length of an ellipse with eccentricity, ε, and for that reason its integral is called the Elliptic Integral of the Second Kind, given by Eqn. 3-37 and graphed in Figure 3-7.

$$\boxed{\int_0^{\phi}\sqrt{1-\varepsilon^2\sin^2(\theta)}\;\mathrm{d}\theta \qquad \varepsilon\le 1} \qquad \text{Eqn. 3-37}$$

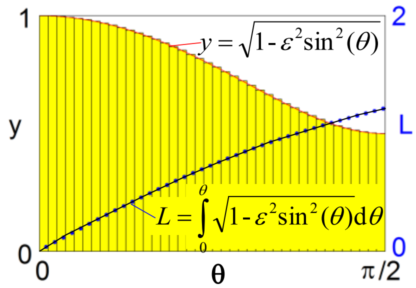


Figure 3-7 $[1-\varepsilon^2\sin^2(\theta)]^{½}$ and the Elliptic Integral.

Note: There are two simple cases when you can integrate Eqn. 3-37 exactly. When ε = 0, the ellipse rounds to a circle and when ε = 1 it flattens to a line.

Three other integrands whose integrals do not have analytic expressions are x^x, sin(x)/x and e^x/x. The last two also exhibit improper behavior (mathematically speaking) and that is why their integrals are called...

3.13 Improper Integrals

Integrals that have 1: unbounded domains or, 2: integrands that become infinite, or 3: have denominators that equal 0 somewhere in the domain are called improper integrals. Despite this, some improper integrals converge. For example, since sin(x) ≈ x as $x \to 0$, the integrand of sin(x)/x is finite and can be calculated. The integrand, $(1-x)^{-0.5}$, (red curve in Figure 3-8) becomes infinite at x = 1, but its integral over the domain, $0 \le x \le 1$ is 2 and is finite. But the integral of e^x/x is not finite when the domain includes x = 0, because $e^x/x \to 1/x$ as $x \to 0$ and the integral of $1/x$ is ln(x), which is infinite at x = 0.

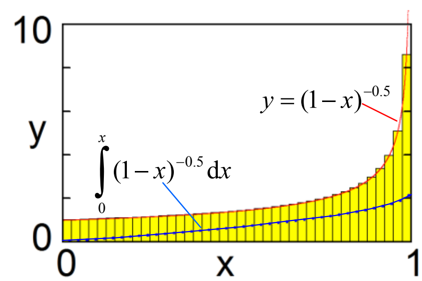


Figure 3-8 $[1-x]^{-½}$ (red) and its finite Improper Integral.

Other improper integrals converge despite infinite bounds. For example, the value of the integral of e^{-kx} over the domain $0 \le x \le \infty$ is $1/k$. And the integral of Eqn. 3-36 over the domain $0 \le x \le \infty$ equals 1 and is therefore also finite.

Improper Integrals present similar convergence problems as do Infinite Series. That makes Series the natural subject of the next chapter.

Conclusion

In this chapter, we have covered so much conceptual ground and have solved so many integrals, including double integrals, that by now you should be seeing double! But, if you have understood most of math in this chapter you are well on Your Royal Road to Genius. So, with these words of well-deserved praise, let's continue in our series of chapters with the Chapter on Series.

CHAPTER 4: SERIES

Introduction

I introduced Infinite Series in Chapter 1 and strongly hinted about their importance. Using just arithmetic, series often are the premier method of approximating difficult functions for calculators, computers and even humans. And we have already seen how closely related series are to both integrals and derivatives.

Two ingenious, useful series derived from Calculus are presented in this Chapter.

Taylor Series express functions as power series (usually of *x*).
Fourier Series express functions as harmonic series of sines and cosines.

Fourier Series are not often included in a first year Calculus course, but they are not too hard at the start and they sure give real good practice and improve understanding of sines and cosines! Another reason I include Fourier Series is that it continues to amaze me that functions which do not even remotely resemble waves can be represented by a series of waves. But let's start with Taylor Series.

4.1 Taylor Series

A Taylor Series expresses a function, f(*x*) as a power series of *x*.

$$f(x) = \sum_{n=0}^{n} a_n x^n \equiv a_0 + a_1 x + a_2 x^2 + \ldots \qquad \text{Eqn. 4-1}$$

As you will see, the coefficient, a_n, is proportional to the n^{th} order derivative of f(*x*).

The problem now is very simple (for a genius) - determine the infinite number of coefficients, a_n. It was Taylor's step of genius to show how easy it is to find all the a_n's. Taylor noted that when $x = 0$ all terms except a_0 are zero. This automatically yields,

$$a_0 = f(0)$$

To find a_1, take the derivative of Eqn. 4-1. Each coefficient is a constant, so its derivative is 0. Thus $\mathrm{d}a_0/\mathrm{d}x = 0$. This leaves us with

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \frac{\mathrm{d}a_0}{\mathrm{d}x} + \frac{\mathrm{d}(a_1 x)}{\mathrm{d}x} + \frac{\mathrm{d}(a_2 x^2)}{\mathrm{d}x} + \frac{\mathrm{d}(a_3 x^3)}{\mathrm{d}x} + \ldots = a_1 + 2a_2 x^1 + 3a_3 x^2 \ldots \equiv \sum_{n=1}^{n} n a_n x^{n-1}$$

Again set $x = 0$ so that all terms except a_1 are zero. This yields the second coefficient.

$$a_1 = \frac{\mathrm{d}f(0)}{\mathrm{d}x}$$

Continue the process. To find a_2, take the second derivative of Eqn. 4-1.

$$\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} = \sum_{n=2}^{n} n(n-1) a_n x^{-2} \equiv 2a_2 + 3\cdot 2a_3 x^1 + 4\cdot 3a_4 x^2 \ldots$$

Once again, set x = 0. And, once again, the series reduces to a single term, namely,

$$a_2 = \frac{1}{2}\frac{\mathrm{d}^2 f(0)}{\mathrm{d}x^2}$$

Do it once more! Take the third derivative and set x = 0. This leaves

$$a_3 = \frac{1}{3\cdot 2}\frac{\mathrm{d}^3 f(0)}{\mathrm{d}x^3}$$

Enough! I hope you can see that generalizing by induction,

$$\boxed{a_n = \frac{1}{n!}\frac{\mathrm{d}^{\mathrm{n}} f(0)}{\mathrm{d}x^n}} \qquad \text{Eqn. 4-2}$$

Recall that $n! = n\times(n-1)\times(\mathrm{n}-2)\times...\times3\times2\times1$ is n factorial. Thus, the Taylor Series is,

$$\boxed{f(x) = \sum_{n=0}^{n} \frac{1}{n!}\frac{\mathrm{d}^{\mathrm{n}} f(0)}{\mathrm{d}x^n} x^n} \qquad \text{Eqn. 4-3}$$

Taylor Series for the Sine and Cosine

The first transcendental function we evaluated using the Taylor Series in Chapter 1 was sin(x), given by Eqn. 1-3,

$$\boxed{\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + ...} \qquad \text{Eqn. 1-3}$$

Let's quickly use Eqn. 4-2 to prove Eqn. 1-3 and then test its accuracy with the first few terms. Calculating the first few values of a_n when f(x) = sin(x) involves little more than taking successively higher order derivatives of sin(x) and evaluating them at x = 0.

$$a_0 = f(x=0) = \sin(0) = 0$$

$$a_1 = \frac{1}{1!}\frac{\mathrm{d}[\sin(x=0)]}{\mathrm{d}x} = \cos(0) = 1$$

$$a_2 = \frac{1}{2!}\frac{\mathrm{d}^2[\sin(x=0)]}{\mathrm{d}x^2} = -\frac{1}{2}\sin(0) = 0$$

$$a_3 = \frac{1}{3!}\frac{\mathrm{d}^3[\sin(x=0)]}{\mathrm{d}x^3} = -\frac{1}{3!}\cos(0) = -\frac{1}{3!}$$

All coefficients with even values of n (and even powers of x) equal 0 because they are some multiple of sin(0) = 0. All coefficients with odd values of n have a magnitude equal to $1/n!$ since cos(0) = 1, but alternate between positive and negative because at alternate steps we must take the derivative of cos(x), which causes a sign change. How do we express alternate values of n? Either $2n$ for even numbers only or $2n + 1$ for odd numbers only. Since it is odd numbers we want for the Taylor Series of sin(x), we write,

$$\sin(x) = \sum_{n=0}^{n} \frac{(-1)^{n+1}}{(2n+1)!} x^{2n+1} \qquad \text{Eqn. 4-4}$$

For the Taylor Series of the cosine, all coefficients with odd values of n (and odd powers of x) drop out (= 0) while the signs of terms with even values of n alternate. You should now try to show that the Taylor Series of cos(x) is, (note: 0! = 1)

$$\cos(x) = \sum_{n=0}^{n} \frac{(-1)^{n}}{(2n)!} x^{2n} \qquad \text{Eqn. 4-5}$$

Truncated Series: Accuracy

No one can calculate an infinite number of terms in a finite lifetime. We have to stop at some point. Infinite series that are cut off after a finite number of terms are truncated (left with only the trunk). By necessity, successive terms in an infinite series that matches a function get progressively smaller. If we are lucky, successive terms get *real* small *real* fast so that the first few terms of the series capture almost all of the function.

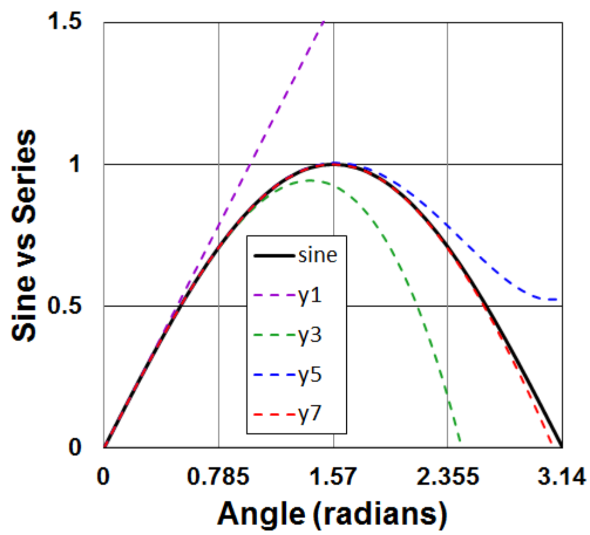


Figure 4-1 Comparing sin(x) with various truncations of the Taylor Series

Let's test our luck and the accuracy of a few truncated versions of Eqn. 4-4. Figure 4-1 displays sin(x) as the black curve in the domain, $0 \le x \le \pi$ and compares it to the Taylor Series, y_1, y_3, y_5, and y_7 that include terms up to x^1, x^3, x^5, and x^7 respectively.

For $x < 0.4$ all approximations are so accurate they are difficult to distinguish from the sine curve. Thus, when x << 1, sin(x) ≈ x. This is the ultimate simplicity. For example (remember that the Radian is the proper unit for angle), when angle, $x = 15° = \pi/12 \approx 0.261799$, using the approximation, $\boldsymbol{x \approx \sin(x)} = \sin(\pi/12) \approx 0.258819$ causes an error of only 1.1%. But as x increases, we need to include more and more terms. y_3 deviates noticeably for $x > 1$ and y_5 for $x > 1.5$. Only y_7 remains reasonably accurate as $x \to \pi$. Even so, this is quite impressive especially considering that y_7 only includes the first 4 terms of an infinite series.

Taylor Series for the Exponential

The Taylor Series for f(x) = e^x is simple because the derivative of e^x is e^x. Furthermore, at $x = 0$ all coefficients involve the factor $e^0 = 1$.

$$a_0 = f(x=0) \;= e^0 = 1$$

$$a_1 = \frac{1}{1!}\frac{\mathrm{d}e^{(x=0)}}{\mathrm{d}x} \quad = e^0 = 1$$

$$a_2 = \frac{1}{2!}\frac{\mathrm{d}^2 e^{(x=0)}}{\mathrm{d}x^2} = \frac{1}{2!}e^0 = \frac{1}{2!}$$

$$a_3 = \frac{1}{3!}\frac{\mathrm{d}^3 e^{(x=0)}}{\mathrm{d}x^3} = \frac{1}{3!}e^0 = \frac{1}{3!}$$

By induction, a_n=1/n!, so that the Taylor Series for e^x is,

$$\boxed{e^x = \sum_{n=0}^{n} \frac{1}{n!}x^n} \qquad \text{Eqn. 4-6}$$

Table 4-1 shows how the percent error of the Taylor Series approximation of e^x when x = 1, 2, and 5 decreases as the number, n of terms increases. What you see are general results. First, the more terms, the smaller the error. Second, the larger x is, the more terms must be included to make the error smaller than a given percentage.

n	Taylor e^1	% error	Taylor e^2	% error	Taylor e^5	% error
0	1	63.21	1	86.47	1	99.33
1	2	26.42	3	59.40	6	95.96
2	2.5	8.03	5	32.33	18.5	87.53
3	2.6667	1.90	6.33	14.29	39.3	73.50
4	2.7083	0.37	7	5.27	65.4	55.95
5	2.7167	0.06	7.27	1.66	91.4	38.40
-----	------------	------------	------------	------------	------------	------------
10	2.7183	1.0E-06	7.39	8.3E-04	146.4	1.37

Table 4-1 The dependence of accuracy of the Taylor Series for e^1, e^2, and e^5 on the number, n of terms included.

4.2 Getting Complex: Imaginary, Real and Complex Numbers

If, by any chance, you noticed that aside from those pesky alternating + and – signs, the Taylor series for sin(x) (Eqn. 4-4) and cos(x) (Eqn. 4-5) match the odd and even terms of the series for e^x, (Eqn. 4-6) then you are definitely morphing into a genius.

What is the necessary flash of insight that will enable us to combine sin(x) and cos(x) to get e^x? You will need quite an imagination because you will have to know about imaginary numbers.

The insight starts by noting that the sign changes every 2nd term in Eqns. 4-4 and 4-5. Since each successive term is a multiple or higher power of the previous term it makes sense that the sin and cosine series require a number whose square equals -1.

$$-1 = i^2$$

where i stands for imaginary. Taking the square root of each side we find that

$$\boxed{i = \sqrt{-1}} \qquad \text{Eqn. 4-7}$$

Everyone uses the symbol, i to represent the basic, unit magnitude imaginary number. You have surely seen imaginary numbers before. Even so they really test our conceptual abilities. Most numbers we deal with are real numbers. Positive integers are easiest to conceive. I have 3 chocolate bars. Positive fractions are a bit more difficult. If I have to share my 3 chocolate bars equally with 6 people (including myself) I get only 2 chocolate bars. Only kidding! Each of us gets ½ of a chocolate bar. Irrational numbers are somewhat harder. They are numbers that are not a ratio of any integers. When you smash a cold chocolate bar, it shatters into various pieces that are sure to be irrational fractions of the entire bar. The world is full of irrational numbers. Famous irrational numbers include $\sqrt{2}$, π, e, and the golden ratio, $\phi \approx 1.618$ [defined by $\phi = 1/(\phi - 1)$].

Negative numbers are also more difficult to comprehend until I realize that I have no assets and owe the bank $53,570. (This makes my net value -$53,570.) And even though mathematicians and philosophers have had such difficulty with 0, the singer, Tom Jones can quickly teach us about 0 – "I, I who have nothing". Come to think of it, even before you began this book your knowledge and understanding of numbers was extremely highly sophisticated by historical standards.

But imaginary numbers, what can they be? Clearly, they are linked somehow to real numbers since $i^2 = -1$ and $i^4 = +1$. You might well think that they only exist in the strange realm of math as solutions to troublesome equations. But some of those troublesome equations arise in physics, and the imaginary numbers are involved in solutions that have real, physical meaning. I'll wave you off for now but soon give a few examples.

First though, we combine real and imaginary numbers. You may find that to be both simple and complex at the same time! (That was a pun!)

Complex Numbers: Combining Real and Imaginary Numbers

We humans are complex. We are forced to live in the real world, deal with real facts, go to a real school and, earn a real living (even if we parasitically live off the work of others). At the same time our imaginations never let go. Our thoughts wander, we daydream by day and dream by night. We create stories, works of art, music or even equations. Some of our creations turn out to be real – there is a link between the world of imagination and the world of reality. So too is the case with real and imaginary numbers.

A complex number is a combination of a real number and an imaginary number, such as $3 + 4i$. Let's add and multiply two complex numbers $(a + ib)$ and $(c + id)$ and see what happens. The four letters, a, b, c, and d all represent real numbers. Be sure to group real and imaginary parts and notice that $ib \times id = i^2bd = -bd$ (which is real).

$$(a+ib)+(c+idx)=a+b+i(c+d)$$
$$(a+ib)\times(c+idx)=ac-bd+i(ad+bc)$$

If two numbers are complex, it is possible that their sum or product can be pure real or pure imaginary. The second equation above implies that the product of two complex numbers is,

1. pure real if $ad = -bc$
2. pure imaginary if $ac = +bd$.

Every complex number, $x + iy$, has a mirror image, $x - iy$ called its **complex conjugate**. Two numbers are complex conjugates when their real parts are equal but their imaginary parts have equal magnitude but opposite sign.

Both the sum and the product of a number and its complex conjugate are real. Eqn. 4-8 proves this for the product.

$$(x+iy)(x-iy)=x^2+y^2=r^2 \qquad \text{Eqn. 4-8}$$

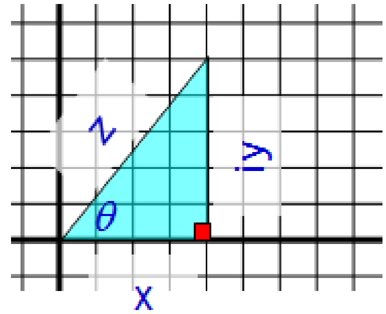


Figure 4-2 Cartesian view of complex numbers

Gauss (that omnipresent math genius again) found a compact and visual way to represent complex numbers. He took a page or rather a graph right out of Cartesian Geometry. He represented real numbers as the *x* axis and imaginary numbers as the *y* axis, as in Figure 4-2. The rest of the graph – the entire area except the axes – consists of complex numbers, with real and imaginary parts. Using Cartesian geometry for complex numbers also correctly suggests that the Pythagorean Theorem applies to complex numbers. The analogy is that the real and imaginary parts of a complex number represent the sides of a right triangle while the complex number itself represents the hypotenuse. Thus, if we call the complex number, *z* then

$$\boxed{z=x+iy} \qquad \text{Eqn. 4-9}$$

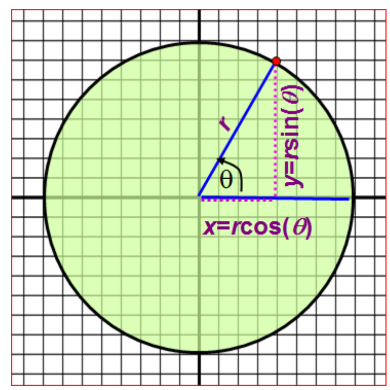


Figure 0-16 Cartesian view of complex numbers in terms of r and θ.

Figure 0-16 (repeated here) and the Pythagorean Theorem imply that 1: the magnitude or absolute value of *z*, is,

$$|z|=r=\sqrt{x^2+y^2}$$

2: we can write Eqn. 4-9 as

$$\boxed{z=r[\cos(\theta)+i\sin(\theta)]} \qquad \text{Eqn. 4-10}$$

Thus, any complex number can be expressed in terms of a magnitude or radius, *r* and an angle, θ. Next, let's show that…

Sines and Cosines are Exponentials! Wow!

We got complex. Now, let's get brave! We can't equate the Taylor Series for sin(θ) and cos(θ) with the Taylor Series for e^{θ}, But, we can with the Taylor Series for $e^{i\theta}$! Just consider the higher order derivatives of $e^{i\theta}$ (and remember that $i^2 = -1$: $i^3 = -i$: $i^4 = 1$….)!

$$\frac{\mathrm{d}e^{i\theta}}{\mathrm{d}x}=ie^{i\theta} \qquad \frac{\mathrm{d}^2e^{i\theta}}{\mathrm{d}x^2}=-e^{i\theta} \qquad \frac{\mathrm{d}^3e^{i\theta}}{\mathrm{d}x^3}=-ie^{i\theta} \qquad \frac{\mathrm{d}^4e^{i\theta}}{\mathrm{d}x^4}=e^{i\theta}$$

Thus, the Taylor Series for $e^{i\theta}$ is,

$$e^{i\theta}=\sum_{n=0}^{n}\frac{i^n}{n!}\theta^n=1+i\theta-\frac{1}{2!}\theta^2-\frac{i}{3!}\theta^3+\frac{1}{4!}\theta^4+\frac{i}{5!}\theta^5-.....$$

Lo and behold, real and imaginary terms alternate. And if we separate all real terms from all imaginary terms, we get two series - one for cos(θ) and one for *i*·sin(θ)!

$$e^{i\theta} = \left(1 - \frac{1}{2!}\theta^2 + \frac{1}{4!}\theta^4 - \ldots\right) + i\left(\theta - \frac{1}{3!}\theta^3 + \frac{1}{5!}\theta^5 - \ldots\right)$$
$$= \quad (\cos(\theta)) \quad + \quad i(\sin(\theta))$$

Wow! Let me repeat this fundamental, unexpected, and amazing result.

$$\boxed{e^{i\theta} = \cos(\theta) + i\sin(\theta)} \qquad \text{Eqn. 4-11}$$

Substituting Eqn. 4-11 into Eqn. 4-10 leads directly to

D'Moivre's Formula

$$\boxed{z = x + iy = r\big(\cos(\theta) + i\sin(\theta)\big) = re^{i\theta}} \qquad \text{Eqn. 4-12}$$

D'Moivre's formula is famous among mathematicians. It means that every complex number consists of a magnitude, *r*, and a direction (or phase), θ. It implies the strange but fundamental result that multiplying complex numbers amounts to multiplying their magnitudes and adding their angles. (Angle θ is an exponent. Remember that multiplying numbers is equivalent to adding their exponents.)

Changing Sines (to Exponentials)

Eqn. 4-11 expresses an exponential in terms of sines and cosines. Let's turn the tables and express sines and cosines in terms of exponentials. Watch these tricks. First, add ½$e^{i\theta}$ and its complex conjugate, ½$e^{-i\theta}$. This becomes cos(θ).

$$\frac{1}{2}\left[e^{i\theta} + e^{-i\theta}\right] = \frac{1}{2}\left[\cos(\theta) + i\sin(\theta) + \big(\cos(\theta) - i\sin(\theta)\big)\right] = \cos(\theta)$$

Next, subtract ½$e^{-i\theta}$ from ½$e^{i\theta}$ and divide each term by *i*. This becomes sin(θ).

$$\frac{1}{2i}\left[e^{i\theta} - e^{-i\theta}\right] = \frac{1}{2i}\left[\cos(\theta) + i\sin(\theta) - \big(\cos(\theta) - i\sin(\theta)\big)\right] = \sin(\theta)$$

Repeating, but putting the cosine and sine on the left hand side yields,

$$\boxed{\cos(\theta) = \frac{1}{2}\left[e^{i\theta} + e^{-i\theta}\right]} \qquad \text{Eqn. 4-13}$$

$$\boxed{\sin(\theta) = \frac{1}{2i}\left[e^{i\theta} - e^{-i\theta}\right]} \qquad \text{Eqn. 4-14}$$

I bet you can combine Eqn. 4-13 and Eqn. 4-14 to express tan(θ) in terms of exponentials. (Hint: tan = sin/cos).

Hyperbolic (Unimaginary) Trigonometric Functions are a Cinch

Eqns. 4-13 and 4-14, which express sines and cosines in terms of exponentials, $e^{i\theta}$ and $e^{-i\theta}$, got mathematicians so excited they took the same combinations without the *i*'s. Forgive me also for using *x* in place of θ. After all, a rose by any other name smells just as sweet, so a variable by any other letter will look just as neat. (That is a poem.) This led to what we call the Hyperbolic Sine (sinh), Cosine (cosh) and Tangent (tanh).

$$\boxed{\sinh(x) = \frac{1}{2}\left[e^{x} - e^{-x}\right]} \qquad \text{Eqn. 4-15}$$

$$\boxed{\cosh(x) = \frac{1}{2}\left[e^{x} + e^{-x}\right]} \qquad \text{Eqn. 4-16}$$

This shows how insensitive mathematicians are. Only they can call something a sinh (pronounced cinch) that is really quite difficult for humans to comprehend.

Hyperbolic and trigonometric functions not only look alike mathematically, they have similar identities and look alike graphically. For example,

$$\boxed{\cosh^2(x) - \sinh^2(x) = 1} \qquad \text{Eqn. 4-17}$$

Eqn. 4-17 explains how the hyperbolic functions got their name. It has the exact same form as Eqn. 0-33 of a hyperbola with *a* = *b* = 1. So, graphing cosh(*x*) as *x* vs. sinh(*x*) as *y* produces a hyperbola as in Figure 0-29. You knew you'd see it again.

Problem: Prove Eqn. 4-17.
Solution: Simply! Subtract the square of Eqn. 4-15 from the square of Eqn. 4-16.

$$\boxed{\frac{1}{4}\left[e^{x} + e^{-x}\right]^2 - \frac{1}{4}\left[e^{x} - e^{-x}\right]^2 = \frac{1}{4}\left\{\left[e^{2x} + 2 + e^{-2x}\right] - \left[e^{2x} - 2 + e^{-2x}\right]\right\} = 1}$$

I will write out the hyperbolic tangent because believe it or not, I have a special feeling for this obnoxious looking function.

$$\boxed{\tanh(x) = \frac{\left[e^{x} - e^{-x}\right]}{\left[e^{x} + e^{-x}\right]}} \qquad \text{Eqn. 4-18}$$

The hyperbolic tangent is graphed with its derivative in Figure 4-3. Tanh(*x*) is a remarkable function. It ranges from -1 to +1 over the domain $-\infty \le x \le +\infty$. Over most of its domain, except around *x* = 0 it increases so slowly that it appears to be horizontal. It is much like a human learning curve. You struggle and struggle and make little progress for hours, days, or weeks. Then, suddenly you make a breakthrough and have a spectacular learning curve (with a large derivative) for a short period, after which you learn little more for the rest of your life. (Hopefully you don't forget too much).

The graphs of the inverse tangent, $2/\pi \tan^{-1}(\pi x/2)$ and its derivative (Figure 4-4) closely resemble the graphs of tanh(*x*) and its derivative (Figure 4-3). When I need a function that resembles tanh(*x*) and $\tan^{-1}(x)$, I prefer to use tanh(x) because it is easy to take derivatives and integrals of exponentials.

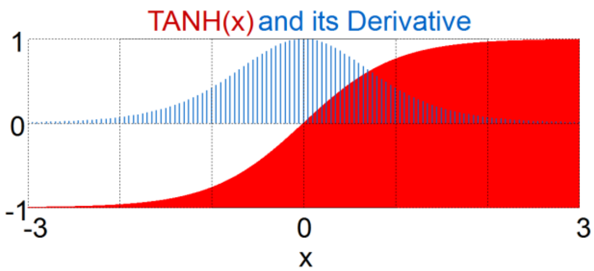


Figure 4-3 Tanh(x) (red curve) and its derivative (blue lines).

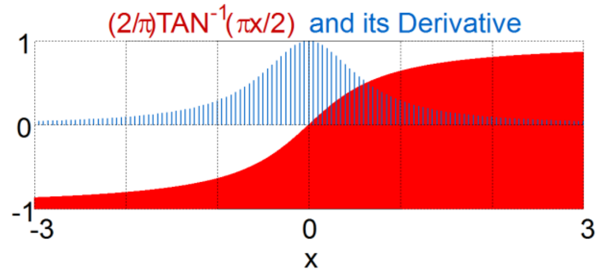


Figure 4-4 $2/\pi\tan^{-1}(\pi x/2)$ (red curve) and its derivative (blue lines) resembles tanh(x)

Each hyperbolic function has almost the same relationship to its derivative as the corresponding trig function does, except perhaps for a minus sign. Working this out for sinh(x) shows it is a cinch. Thus,

$$\boxed{\frac{\mathrm{d}[\sinh(x)]}{\mathrm{d}x}=\frac{1}{2}\frac{\mathrm{d}\left[e^{x}-e^{-x}\right]}{\mathrm{d}x}=\frac{1}{2}\left[e^{x}-(-e^{-x})\right]=\cosh(x)} \quad \text{Eqn. 4-19}$$

You can work out the derivatives of cosh(x) and tanh(x). They are,

$$\boxed{\frac{\mathrm{d}[\cosh(x)]}{\mathrm{d}x}=+\sinh(x)} \quad \text{Eqn. 4-20}$$

$$\boxed{\frac{\mathrm{d}[\tanh(x)]}{\mathrm{d}x}=1-\tanh^{2}(x)=\frac{1}{\cosh^{2}(x)}} \quad \text{Eqn. 4-21}$$

Taylor Series for ln(1 + *x*)

Read the last part of the heading once again! Does the (1 + x) trouble you? It really bothered me at first when I learned Calculus. Why ln(1 + x)? Why not simply ln(x)?

Let's learn the hard way and do it. As we do it, you must remember that 1: d[ln(x)]/dx = 1/x and, 2: all derivatives must be evaluated at x = 0

$$\ln(x)=\sum_{n=0}^{n}\frac{1}{n!}\frac{\mathrm{d}^{n}\ln(0)}{\mathrm{d}x^{n}}x^{n}=\ln(0)+\frac{1}{1!}\left(\frac{1}{0}\right)x^{1}+\frac{1}{2!}\left(\frac{-1}{0^{2}}\right)x^{2}+\ldots$$

Do you should see the problem? No matter what the value of x, every single term in the series is infinite because ln(0) = $-\infty$, and all orders of all the derivatives of ln(x) $\propto x^{-n}$ ($\propto$ means proportional to) are infinite when x = 0. Seeing this problem, mathematicians immediately gave up and squirmed their way to something that worked. (Expand the series starting from 1, not 0.) Discretion is the better part of valor!

The mathematicians said, "let's try writing a series for ln(1 + x) since ln(1 + x) equals 0 when x = 0 and since all its derivatives, starting with d[ln(1 + x)]/dx = 1/(1 + x) are finite when x = 0." So here is the equation.

$$\ln(1+x) = \sum_{n=0}^{n} \frac{1}{n!}\frac{\mathrm{d}^n \ln(1)}{\mathrm{d}x^n} x^n = \ln(0) + \frac{1}{1!}\left(\frac{1}{1}\right)x^1 + \frac{1}{2!}\left(\frac{-1}{1^2}\right)x^2 + \frac{1}{3!}\left(\frac{-1(-2)}{1^3}\right)x^3 + \ldots$$

Simplifying leads to the Taylor Series,

$$\boxed{\ln(1+x) = \frac{x^1}{1} - \frac{x^2}{2} + \frac{x^3}{3} - \ldots.} \qquad \text{Eqn. 4-22}$$

4.3 Some Infinite Series <u>are</u> Infinite: A Question of Convergence

There is one tiny trouble with Eqn. 4-22. It only works when -1 < x < 1. Outside this domain (|x| ≥ 1) the series fails dismally. In fact it blows up or diverges. This means the more terms you add the larger the magnitude of the series even though ln(1 + x) itself is finite for all values of x > 1. Ugh!

Figure 4-5 compares ln(1 + x) with three truncations (2, 6, and 8 terms) of Eqn. 4-22. It gives a hint of the bad behavior of the series when x > 1. For all |x| < 1 the more terms included in the series the more accurate it is. That's good! But for |x| > 1, the more terms included in the Taylor Series expansion, the greater its discrepancy with ln(1 + x). That's bad!

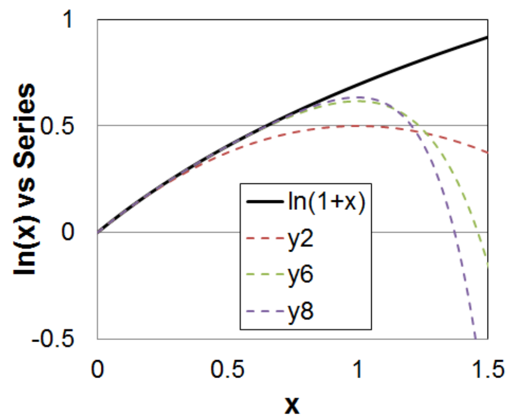


Figure 4-5 Comparing ln(1+x) with 3 truncations of the Taylor Series

This brings us to the nasty but necessary job of determining when and if a series converges or diverges. Let's gather our knowledge. Naturally, if successive terms get larger, any infinite series will diverge. Geometric series converge so long as each successive term is smaller in magnitude. So a Taylor Series will converge if successive terms are smaller than the successive terms of a geometric series. I have covered my bases in making this statement and there is a standard name for this criterion, called…

The Ratio Test. If the ratio, r, of successive terms in an infinite series remains smaller in magnitude than a definite number, R < 1 then the series will converge. By contrast, if |r| ≥ 1 then the series diverges.

The reason to state the Ratio Test so carefully is illustrated by the two infinite series below. They might *seem* to look and behave alike but $S_{1/n}$ diverges and S_{1/n^2} converges.

$$S_{1/n} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \ldots.\frac{1}{n} + \ldots.$$

$$S_{1/n^2} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \ldots.\frac{1}{n^2} + \ldots.$$

For both of these series the ratios of successive terms, $[n/(n+1)]$ and $[n/(n+1)]^2$ are less than 1 but approach 1 as $n \to \infty$. Thus, they are not confined within a geometric series by a ratio $r < R < 1$. For these more pesky series there is the Integral Test.

The Integral Test. If there is an integral that the series corresponds to, the series will converge if the integral converges and diverge if the integral diverges.

The integral test works because series are, like Riemann sums, finite difference versions of integrals. Comparing the two series above to integrals proves that $S_{1/n}$ is infinite and hence diverges,

$$S_{1/n} = \sum_{i=1}^{\infty} \frac{1}{i} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + ... \approx \int_1^{\infty} \frac{\mathrm{d}x}{x} = \ln(\infty) - \ln(1) = \infty \qquad \text{Eqn. 4-23}$$

By contrast, S_{1/n^2} is finite and hence, converges.

$$S_{1/n^2} = \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + ... \approx \int_1^{\infty} \frac{\mathrm{d}x}{x^2} = -\frac{1}{\infty} - \left(-\frac{1}{1}\right) = 1 \qquad \text{Eqn. 4-24}$$

Remember that the integrals only approximate the Series. They won't mislead you about a series with an infinite sum, but will not give an accurate value for a series with a finite sum. For, example the sum for the series of Eqn. 4-24 is $\pi^2/6 \approx 1.644...$ (not 1.0) and you can do it with Fourier Series although the proof is almost as long as the series!

L'Hôpital's Rule: A Singular Twist on Limits

One of the great applications of Taylor Series was derived by Johann Bernoulli but through a quirk of history is universally called L'Hôpital's Rule. The story is almost too juicy to believe, but let's look at the rule first.

Imagine two functions, $u(x)$ and $v(x)$ that both approach 0 as $x \to 0$. At $x = 0$, the fraction, $u(0)/v(0) = 0/0$ is indeterminate. Taylor Series show that the limit of the ratio, $u(x)/v(x)$ as $x \to 0$ is equal to the ratio of their derivatives at $x = 0$.

$$\boxed{\lim_{x\to 0}\left(\frac{u(x)}{v(x)}\right) \approx \frac{u(0) + \frac{\mathrm{d}}{\mathrm{d}x}[u(0)]x}{v(0) + \frac{\mathrm{d}}{\mathrm{d}x}[v(0)]x} = \frac{\frac{\mathrm{d}}{\mathrm{d}x}[u(0)]}{\frac{\mathrm{d}}{\mathrm{d}x}[v(0)]}}$$

Thus, for example, if $u(x) = x$ and $v(x) = \sin(x)$ then Eqn. 1-3 and L'Hôpital's Rule tell us that the limit of $x/\sin(x) = 1$ as $x \to 0$.

Now for the juicy story! L'Hôpital was a rich baron who fell in love with Calculus. He paid Bernoulli both as a tutor and for his discoveries. L'Hôpital then wrote the first text in differential Calculus. In it, he borrowed and acknowledged Bernoulli's discovery - what we now call L'Hôpital's Rule. But Johann Bernoulli was a very jealous man, and as time passed, he not only claimed credit for his discovery, but also for L'Hôpital's text. Johann was so jealous he even attempted (unsuccessfully) to stifle the mathematical genius of his own son, Daniel. So the misnaming of his rule is almost a form of cosmic revenge!

4.4 Fourier Series: The Heat's On!

Fourier Series express any function in terms of the harmonics or frequencies of sine and cosine waves. Harmonics are shorter waves that repeat an integral number of times in the same space or time that the fundamental wave has a single cycle, as in Figure 4-6 where cos(x), cos($3x$) and cos($9x$) are all shown. Wave amplitude (half the height) is a quantity that is independent of wavelength but for aesthetic reasons, the higher harmonics in Figure 4-6 are depicted with smaller amplitude than the longer waves.

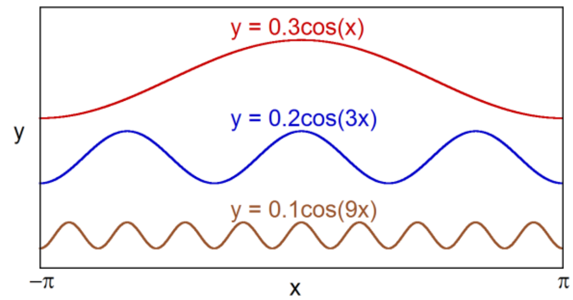


Figure 4-6 A wave with frequency 1 and harmonics with frequencies 3 and 9.

More complicated Fourier *Integrals* can be used to find overtones, all higher frequencies (not just integer multiples) of the fundamental tone. Our voices and all musical instruments produce sound waves with a range of frequencies. Figure 4-7 compares the complex waveform of a violin and piano with a sine wave. Using Fourier Analysis a Synthesizer can reproduce the harmonics and overtones, and therefore the sound of each instrument.

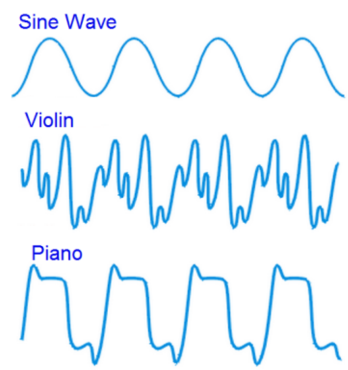


Figure 4-7 Pressure waveform of a sine wave vs a note from a violin and piano.

Aside from the great importance of Fourier Series in mathematics and many branches of science, it will give you excellent practice working with sines and cosines, which is one reason I have included it. (I also wanted to relearn Fourier Series!) To say that I didn't like Fourier Series as a student would be a gross understatement, even though I recognized it was good math medicine. Unfortunately for me, I didn't take the full dose of Fourier Series and Integrals for years, so my problem solving skills suffered accordingly. If you behave more nobly than I did, you will reap great rewards.

The idea that any function can be expressed as a series of sine and cosine waves is a combination of the profound and the profoundly lucky. It has ancient roots based on astronomy. When the Ancients watched Mars, Jupiter and Saturn, they noticed that for most of the year, they appeared earlier each succeeding night and a little further west of the stars, but for a small part of the year, they reversed direction, appearing later each night and further east relative to the stars. The alternating forward and backward motion was puzzling but the ancient Babylonians became such skilled sky watchers that they could predict the puzzling times and locations of the planets in the sky long in advance.

The Ancient Greeks knew Babylonian Astronomy and tried to improve it. Most, except for Aristarchus, assumed that all celestial objects circle the Earth, which they naturally (through no ego at all) placed at the center of the Universe. They also insisted that all planets orbit the Earth in circles. But circular orbits go in one direction only.

The only way for the Ancients to stubbornly cling to these two fixed ideas was to superimpose a small circle or epicycle on the orbit's grand circle. It took more than 1500 years before Copernicus finally tossed epicycles in the trash where they belonged.

I can't say if the idea of epicycles led directly to the concept that complex functions can be built up of simple waves, but it is quite possible. Beginning around 1750, mathematicians led by Daniel Bernoulli (son of the jealous Johann) and Leonhard Euler used simple trigonometric series to solve some astronomical problems. Finally, the complete series was worked out by Jean Baptiste Joseph Fourier by 1822 when he solved the problem of how heat is conducted through solids. (Fourier did other work on heat – he was the first to propose that the Earth is warmed by the Greenhouse Effect.)

The number of significant applications of Fourier Series is mind boggling. All phenomena that involve waves use Fourier Series. All aspects of telecommunications, including the phone, radio and TV use Fourier Series because they involve transmitting and receiving electromagnetic waves that must be converted to and from sound waves.

A Fourier Series expresses a function, f(x) as a series of sine and cosine wave harmonics within a specified domain for x (**or time, *t***). The math is simplest if the domain is, $-\pi \le x \le \pi$. This domain fits complete waves for cos(nx) and sin(nx).

$$\boxed{f(x) = \sum_{n=0}^{\infty} \left[a_n \cos(nx) + b_n \sin(nx) \right] \qquad -\pi \le x \le \pi} \qquad \text{Eqn. 4-25}$$

Just as with Taylor Series, the job is to find the constant coefficients, a_n and b_n.

I want to repeat that the Fourier Series specifies a domain for x (or t). The series is valid in that domain but almost surely not valid outside it because while the series is periodic, most functions such as $y = x$ or $y = x^2$ are not periodic.

I also want to repeat that it is simplest to use the domain, $-\pi \le x \le \pi$. In that domain cos(x) and sin(x) go through one complete cycle, cos($2x$) and sin($2x$) go through two complete cycles, and cos(nx) and sin(nx) go through n complete cycles. The domain is centered at 0. That will make it easier to find the constant coefficients, a_n and b_n.

With Taylor Series the step of genius to find the coefficients is to take successive derivatives and evaluate the series at a fixed value of x, usually $x = 0$. By doing this, for each derivative all terms but one drop out. With Fourier Series two steps of genius are needed to find the coefficients.

Step #1: Multiply each term in the series by sin(mx) or cos(mx), where m is any integer,
Step #2: Integrate over the domain. As with the Taylor Series, this eliminates all but one term of the series. To begin, integrate each term in Eqn. 4-25 over the domain of x

$$\int_{-\pi}^{\pi} f(x)dx = \int_{-\pi}^{\pi} \sum_{n=0}^{\infty} \left[a_n \cos(nx) + b_n \sin(nx) \right] dx \qquad \text{Eqn. 4-26}$$

Integrating zillions of terms in Eqn. 4-26 may seem impossible, but by luck is relatively easy. Every integral on the right hand side except for $a_0\cos(0x) = a_0$ involves a sine or cosine times a constant. But the integral of a sine or cosine over a complete cycle is 0 because the negative and positive areas cancel (see Figure 4-8).

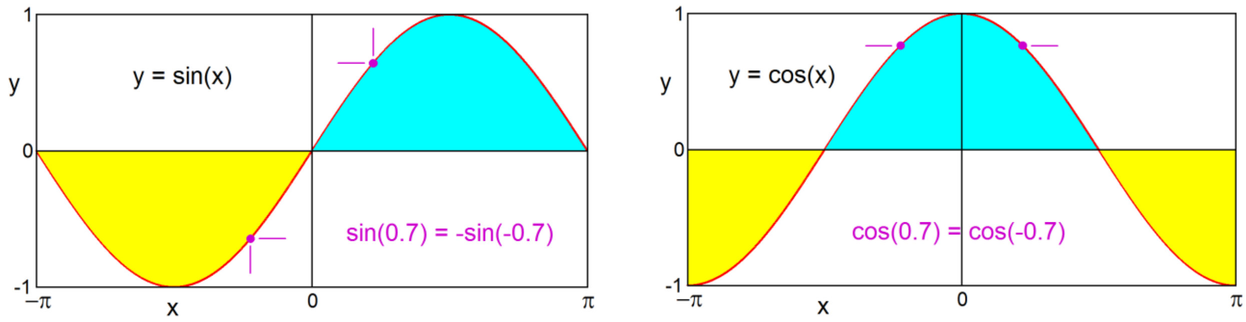


Figure 4-8 The integral of any wave over a complete cycle equals 0 because positive (blue) and negative (yellow) areas are equal and cancel.

This leaves only a_0, which is given by...

$$\int_{-\pi}^{\pi} f(x)dx = 2\pi a_0 \qquad \text{Eqn. 4-27}$$

Remember! Even if f(x) is so tough that you can't find an analytic form of its integral, you can always integrate numerically using a technique such as Riemann Sums.

Next, multiply every term in the series by cos(mx), where m is any positive integer and integrate over the domain. We must prove that n = m is the only term that is not 0.

$$\int_{-\pi}^{\pi} \cos(mx) f(x)dx = \int_{-\pi}^{\pi} a_m \cos^2(mx)dx = \pi a_m \qquad \text{Eqn. 4-28}$$

To do this we must prove that every integral for $m \neq n$ in Eqn. 4-25 equals zero, i. e.,

$$\int_{-\pi}^{\pi} \cos(mx)\cos(nx)dx = 0 \qquad \int_{-\pi}^{\pi} \cos(mx)\sin(nx)dx = 0 \qquad \text{Eqn. 4-29}$$

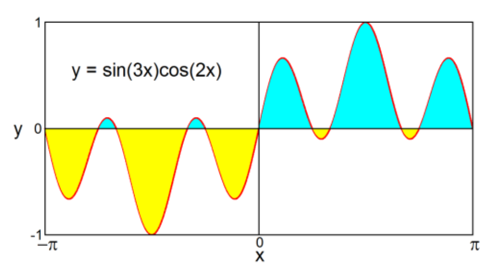


Figure 4-9 Positive (blue) and negative areas (yellow) of any antisymmetric function cancel.

It is easy to prove that the second integral in Eqn. 4-29 is zero using reasoning about symmetry. Cos(x) is symmetric about x = 0 because +cos(-x) = +cos(x). By contrast, sin(x) is antisymmetric about x = 0 because +sin(-x) = -sin(x). The product of a symmetric function and an antisymmetric function is antisymmetric, while the product of two symmetric functions or two antisymmetric functions is symmetric.

The integral or area of any antisymmetric function is zero when the domain of x is centered at the axis of symmetry (here, x = 0) because the negative area is exactly equal to the positive area. Symmetry is visual (like Rorschach ink blots), so if you don't understand the reasoning, simply look at Figure 4-9, which graphs y =

sin(3x)·cos(2x). It should convince you that the net area of an antisymmetric function is zero because every hill of positive area (blue) is cancelled exactly by the every valley of negative area (yellow).

To prove the first integral of Eqn. 4-29 equals zero, use the multiple angle formula, Eqn. 0-13a to show that its integrand equals ½{cos[(m + n)x] + cos[(m - n)x]}.

$$\cos(mx)\cos(nx) = \frac{\cos[(m+n)x]+\cos[(m-n)x]}{2} \qquad m \neq n \qquad \text{Eqn. 4-30}$$

Eqn. 4-30 proves that the integral, ∫cos(mx)·cos(nx)dx = 0 over a complete cycle for all $m \neq n$ because it is equal to the integral of two simple cosine waves, each of which is 0 over a complete cycle. But when $m = n$, cos[(m - n)x] = cos[0] = 1, so Eqn. 4-30 and Eqn. 3-18 prove that the integral of $\cos^2(nx)$ from $-\pi \leq x \leq \pi$ is the same as the integral of ½ from $-\pi \leq x \leq \pi$. So, we have just proven that every integral for $m \neq n$ in Eqn. 4-26 equals zero, and only the integral for $m = n$ remains. Thus, the equation for a_n,

$$\int_{-\pi}^{\pi} f(x)\cos(nx)dx = \int_{-\pi}^{\pi} a_n \cos^2(nx)dx = \pi a_n \qquad \text{Eqn. 4-31}$$

You should go through this whole painful process again to find the equations for the constant coefficients, b_n (Just multiply the series by sin(mx) and integrate) because it is good practice, but I, as author, will only give you the final results for both a_n and b_n.

$$\boxed{a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx \qquad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx \quad n \geq 1} \qquad \text{Eqn. 4-32a}$$

$$\boxed{b_0 = 0 \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx \quad n \geq 1} \qquad \text{Eqn. 4-32b}$$

Now that we finally have the coefficients for the Fourier Series, let's see how closely Fourier Series can approximate two simple functions, $y = x$ and $y = x^2$. To save time, because all mathematicians are lazy, I will play the even and odd game again.

Problem: Find the Fourier Series for $y = x$.

Solution: Since x is antisymmetric, so is the product, x·cos(nx) for any n. Therefore, every one of their integrals over a domain centered at $x = 0$ are zero because – you got it – the positive and negative areas cancel and thus, every $a_n = 0$.

Therefore, we only have to find b_n. This is a job that calls for integration by parts

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x\sin(nx)\mathrm{d}x = -\frac{1}{n\pi}\int_{-\pi}^{\pi} x\mathrm{d}[\cos(nx)] =$$

$$b_n = -\frac{1}{n\pi}\left[\pi\cos(n\pi) - (-\pi)\cos(n\pi)\right] + \frac{1}{n\pi}\int_{-\pi}^{\pi}\cos(nx)\mathrm{d}x$$

The integral of cos(nx)dx is 0 because the negative and positive parts of any cosine over a complete cycle cancel because they are equal. This leaves b_n = -2cos($n\pi$)/n. When n is even, cos($n\pi$) = 1, and when n is odd, cos($n\pi$) = -1. The final result, then is

$$b_n = \frac{2}{n}(-1)^{n+1}$$

So, now we write x using Fourier Series as,

$$x = \frac{2}{1}\sin(x) - \frac{2}{2}\sin(2x) + \frac{2}{3}\sin(3x) - \frac{2}{4}\sin(4x) + ... \quad -\pi \le x \le \pi$$

This sure is a long expression for a simple function like $y = x$! Figure 4-10 shows that at least 20 terms are needed for the Fourier Series to produce a close approximation over most of the domain, but no matter how many terms are included, the series goes wild near the edges of the domain. This is called the Gibbs phenomenon even though Henry Wilbraham discovered and explained it in 1843, 50 years before Willard Gibbs did.

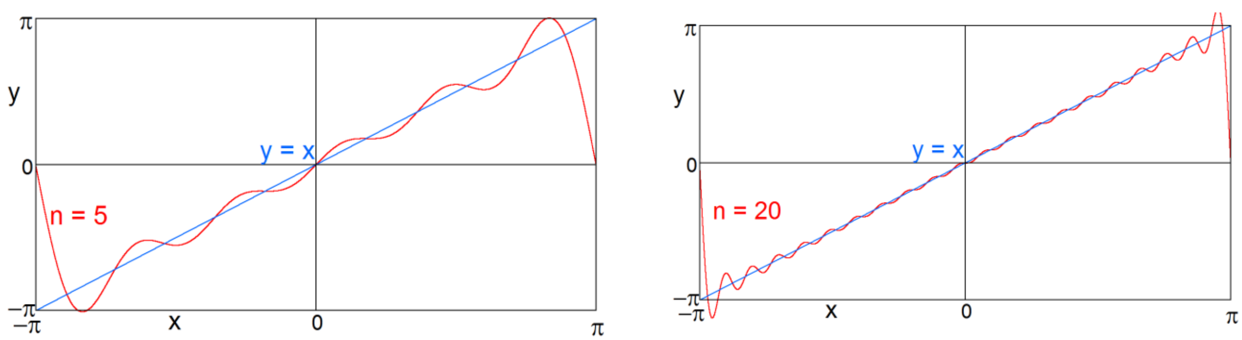


Figure 4-10 Fourier Series approximations for $y = x$ with n = 5 (left) and n = 20 (right).

Problem: Find the Fourier Series for $y = x^2$.
Solution: Since x^2 is symmetric about x = 0, $x^2\cdot\sin(nx)$ is antisymmetric and every b_n = 0, so we only have to calculate the a_n. Calculating a_0 is simplest.

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi}x^2\mathrm{d}x = \frac{\pi^2}{3}$$

Calculating all the other a_n's involves integrating by parts twice since we have to get rid of x^2. This is good practice. The first integration by parts is,

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2\cos(nx)\mathrm{d}x = \frac{1}{n\pi}\int_{-\pi}^{\pi} x^2\mathrm{d}[\sin(nx)] =$$

$$= \frac{1}{n\pi}\left[\pi^2\sin(n\pi) - \pi^2\sin(-n\pi)\right] - \frac{1}{n\pi}\int_{-\pi}^{\pi} 2x\sin(nx)\mathrm{d}x$$

The term in brackets equals 0 because sin($n\pi$) = 0. The second integral by parts is,

$$a_n = -\frac{1}{n\pi}\int_{-\pi}^{\pi} 2x\sin(nx)\mathrm{d}x = \frac{1}{n^2\pi}\int_{-\pi}^{\pi} 2x\mathrm{d}[\cos(nx)] =$$

$$= \frac{1}{n^2\pi}\left[2\pi\cos(n\pi) - 2(-\pi)\cos(-n\pi)\right] - \frac{1}{n^2\pi}\int_{-\pi}^{\pi} 2\cos(nx)\mathrm{d}x$$

The last integral equals 0 because it is a simple cosine over a complete cycle. Since cos($n\pi$) is positive when n is even and negative when n is odd, a_n is given by,

$$a_n = \frac{4}{n^2}(-1)^n \qquad n \geq 1$$

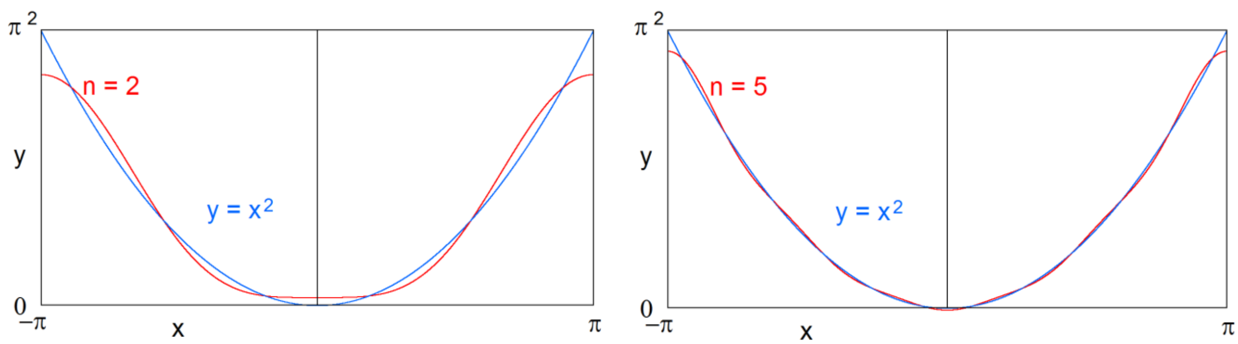


Figure 4-11 Fourier Series approximations (red) for $y = x^2$ (blue) for $n = 2$ (left) and $n = 5$ (right).

Figure 4-11 shows that with just 2 harmonics ($n = 2$) the Fourier Series provides a decent approximation to $y = x^2$, and with 5 harmonics the fit is so good, except near the edges of the domain, that is it hard to tell the difference. This good fit should not be too surprising since the parabola, $y = x^2$ almost looks like a wave around its axis of symmetry. This implies that when a function, such as a telephone or radio signal, consists of a series of superposed waves, Fourier Series will represent it very accurately and help diagnose all its frequencies and their amplitudes.

Fourier Series for Real Data – What to do When You don't Have a Function

What a triumph! Aren't you proud that you can find the Fourier Series of any function when you 1: know the function, 2: define it in the domain from $-\pi \leq x \leq \pi$ and, 3: can integrate it both when it stands alone and when it is multiplied by a sine or cosine.

But just when you start getting smug in math you get hit with a humbling problem. How can you construct the Fourier Series if you don't have a simple, well-defined function, but instead have data given at finite intervals?

For example, consider mean monthly temperatures at San Francisco (SFO) and New York City (NYC). The data is given in Table 4-2 and graphed in Figure 4-12, where the red circles represent the data and the yellow circles, the Fourier Series approximation with only two harmonics.

The annual march of temperature resembles a cosine curve at NYC but is markedly asymmetrical at SFO, rising slowly from January to September and then decreasing rapidly from October to December. If you are surprised at how closely the Fourier Series approximates the real data, perhaps you shouldn't be. After all, the annual cycles of temperature for both cities do resemble waves, so a series composed of waves should be able to match them well.

City	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
SFO	11.2	12.6	13.4	13.9	14.8	15.9	16.4	17.0	17.5	16.9	14.2	11.4
NYC	3.4	3.8	6.1	11.9	17.1	22.2	25.0	24.3	20.2	14.1	9.0	3.3

Table 4-2 Mean monthly *T* at SFO and NYC.

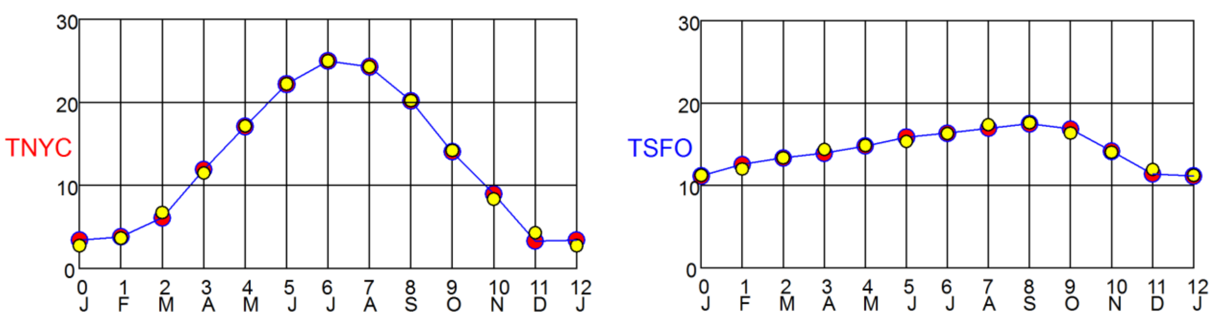


Figure 4-12 Mean monthly temperature at NYC (left) and SFO (right). Red circles and blue lines are the data, yellow circles are the Fourier Series with 2 harmonics.

These excellent results justify all the troubles we may have with finding the Fourier Series for a data set such as temperature. There are three troubles:

1: The temperatures are based on data and are not simple, analytic functions.

2: The data is discontinuous, given at finite (but luckily, equally-spaced) intervals.

3: The domain is not $-\pi \le x \le \pi$ but rather $0 \le x \le 12$.

Troubles #1 and #2 mean that we must use sums instead of integrals to calculate the a_n's and b_n's. Fortunately, the sums have the same format as the integrals for the a_n's and b_n's in Eqn. 4-32a and Eqn. 4-32b. When data is given at finite intervals, the shortest cycle requires 2 data points (alternately up and down), so the maximum number of waves or harmonics in the Fourier Series is 6 when the domain extends from 0 to 12. Trouble #3 means that we must multiply the independent variable, t, by $2\pi/12$ so that sine and cosine waves experience complete cycles in 12 months. With these conditions, the Fourier Series is,

$$\boxed{T(t)=\sum_{n=0}^{6}\left[a_n\cos\left(\frac{2\pi nt}{12}\right)+b_n\left(\frac{2\pi nt}{12}\right)\right]\qquad 0\le t\le 12}\qquad \text{Eqn. 4-33}$$

The coefficients are given by,

$$a_0=\frac{1}{12}\sum_{t=0}^{11}T(t)\qquad a_n=\frac{2}{12}\sum_{t=0}^{11}T(t)\cos\left(\frac{2\pi nt}{12}\right)\qquad \text{Eqn. 4-34}$$

$$b_n=\frac{2}{12}\sum_{t=0}^{11}T(t)\sin\left(\frac{2\pi nt}{12}\right)$$

Watch how I calculate the coefficients using the data for T_{SFO} from Table 4-2 as the example! The coefficient, a_0 is simply the annual average value of T, or,

$$a_0=\frac{1}{12}(11.2+12.6+13.4+...+11.6)=14.6$$

The rest of the coefficients require more work, and are best to do in some program such as EXCEL. An indication of the work needed is shown for a_1 and a_2 for SFO.

Warning: These formulas are long.

$$a_1=\frac{1}{12}\left[11.2\cos\left(\frac{2\pi(0)}{12}\right)+12.6\cos\left(\frac{2\pi(1)}{12}\right)+13.4\cos\left(\frac{2\pi(2)}{12}\right)+...\right]=-2.54$$

$$a_2=\frac{1}{12}\left[11.2\cos\left(\frac{4\pi(0)}{12}\right)+12.6\cos\left(\frac{4\pi(1)}{12}\right)+13.4\cos\left(\frac{4\pi(2)}{12}\right)+...\right]=-0.80$$

After including these and more terms from Eqn. 4-34, and after the b's are calculated in the same manner, the Fourier Series up to the 2nd harmonic for SFO is.

$$T_{SFO}=14.6-2.54\cos\left(\frac{2\pi t}{12}\right)-0.80\cos\left(\frac{4\pi t}{12}\right)-0.99\sin\left(\frac{2\pi t}{12}\right)+0.60\sin\left(\frac{4\pi t}{12}\right)$$

It would be nice to shorten this relatively long equation. That will help us understand it.

The Amplitude and Phase of the Waves of Fourier Series (Harmonic Analysis)

There is a way to shorten the long formulas for the Fourier Series. To do that it helps to repeat Figure 2-2, which shows that cosine and sine waves have the same shape but are out of phase = offset = not aligned. When the crests and troughs of two waves are aligned perfectly, the two waves are in phase. If the cosine curve is moved to the right by phase angle, $\alpha = \pi/2$ it is in phase with (and identical to) the sine curve.

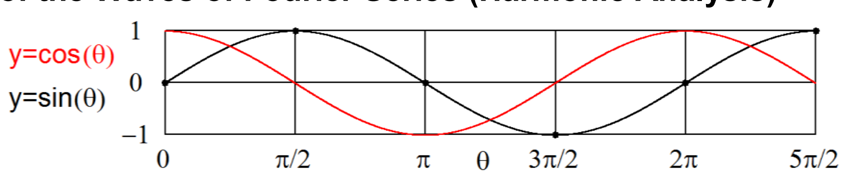


Figure 2-2 Sin waves are Cosine waves moved to the right by phase angle, $\alpha = \pi/2$.

How can we combine the cosine and sine waves for each value of n (each harmonic) into a single cosine wave? Think of a multiple angle formula!

$$a_n \cos(nt) + b_n \sin(nt) = c_n \cos(nt + \alpha_n) = c_n [\cos(nt)\cos(\alpha_n) - \sin(nt)\sin(\alpha_n)]$$

This is really two equations (note the color coding) – one for sin(nt) and one for cos(nt). These equations enable us to calculate both c_n and α_n. When we cancel sin(nt) and cos(nt) – the common terms in each equation, what remains is,

$$\boxed{\begin{aligned} a_n &= c_n \cos(\alpha_n) \\ b_n &= -c_n \sin(\alpha_n) \end{aligned}} \qquad \text{Eqn. 4-35}$$

Note that each final, combined cosine wave has a different amplitude, c_n and phase angle, α_n than the component sine and cosine waves.

To find c_n, square each equation of Eqn. 4-35, add them and use the Pythagorean Theorem. The result is the Equation for the Amplitude,

$$\boxed{c_n = \sqrt{a_n^{\ 2} + b_n^{\ 2}}} \qquad \text{Eqn. 4-36}$$

To solve for α_n, divide b_n by a_n in Eqn. 4-35. Remember that $\sin(\alpha_n)/\cos(\alpha_n) = \tan(\alpha_n)$. The result is the Equation for the Phase

$$\boxed{\tan(\alpha_n) = -\frac{b_n}{a_n} \quad \Rightarrow \alpha_n = \tan^{-1}\left(-\frac{b_n}{a_n}\right)} \qquad \text{Eqn. 4-37}$$

Hooray! Let's apply Eqn. 4-36 and Eqn. 4-37 to the Fourier Series of T's at SFO and NYC. Here are some general interpretations along with specific results from Table 4-3.

1. a_0 is the average value or the mean annual temperature
2. c_1 is the amplitude (half the temperature range) of the annual cycle. It is much greater at NYC than at SFO.
3. c_2 is the amplitude of the semiannual cycle.
5. α_1 tells the day number of maximum temperature of the annual cycle. It is 14 days later for SFO than for NYC.

	SFO	NYC
a_0	14.6	13.3
c_1	2.73	11.22
c_2	1.00	0.63
α_1	201.00	187.00

Table 4-3 Fourier Series constants for monthly T at SFO and NYC.

Moving On

You now know so much about series and waves you should be proud of yourself. But if you want to win the World Series of Waves, you must learn how to make waves and how to make the waves move. In the next chapter we will generate waves (and diagnose and solve other phenomena) when we learn about Ordinary Differential Equations (which have one independent variable). Then, in Chapter 6, we will move waves (and mountains) using Partial Differential Equations (which have more than one independent variable).

CHAPTER 5: ORDINARY DIFFERENTIAL EQUATIONS AND THEIR INTEGRALS

Ordinary Differential Equations (ODE's) are equations that contain derivatives of a single independent variable. Partial Differential Equations (PDE's) (see Chapter 6) contain several independent variables. Because ODE's and PDE's treat rates of change, they involve prediction and crop up in the sciences and in fields such as economics.

Solving Ordinary Differential Equations always amounts to integrating but because so many ingenious techniques are involved, whole courses are dedicated to solving them. Here, we set up and solve some premier examples of ODE's so that you will get a hint of what this important field involves and be able to make your own predictions.

We begin by solving Newton's Equations of Motion for five simple cases. Next, we forecast floods and overdosing. Then we find and solve differential equations* that forecast changes of animal populations, chemical and economic activity, and even tumor sizes. The chapter ends on a chaotic (pun intended as you will see) note by returning to solve finite difference versions of differential equations, as we started doing in Section 1.7.

The five motion cases involve combinations of three forces – gravity, friction, and springs, and provide an excellent introduction to solving Ordinary Differential Equations.

Case #1 is called free fall of an object due to constant acceleration of gravity.
Case #2 is the motion of an object influenced by friction alone.
Case #3 is the motion of a falling object that is also subject to friction.
Case #4 is the motion of a mass attached to a spring without any friction.
Case #5 is the motion of a mass attached to a spring with friction.

5.1 Motion Case #1: Free Fall or Constant Acceleration

We already know the defining equations for acceleration (Eqn. 1-28) and velocity (Eqn. 1-29) and also know that they can be combined by eliminating v to get Eqn. 2-7.

$$\boxed{a=\frac{\mathrm{d}v}{\mathrm{d}t}} \quad \text{Eqn. 1-28} \qquad \boxed{v=\frac{\mathrm{d}s}{\mathrm{d}t}} \quad \text{Eqn. 1-29} \qquad \boxed{a=\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\mathrm{d}s}{\mathrm{d}t}\right]} \quad \text{Eqn. 2-7}$$

The compact result is a second order differential equation for distance, s.

$$\boxed{a=\frac{\mathrm{d}^2 s}{\mathrm{d}t^2}} \qquad \text{Eqn. 5-1}$$

We solved Eqn. 5-1 when acceleration is constant by integrating Eqn. 1-28 and Eqn. 1-29, but did not include the constants of integration. So now let's complete the job by doing the definite integrals. It is standard to start time at t = 0 but since the object may be moving at t = 0, we allow it to have an *i*nitial velocity, v_i.

$$\int_{v_i}^{v}\mathrm{d}v=\int_0^t a\,\mathrm{d}t \quad \Rightarrow \quad v-v_i=at \quad \Rightarrow \quad v=v_i+at \qquad \text{Eqn. 5-2}$$

We also start with distance, s = 0, even though in some problems, such as King Kong

* Whenever we don't specify, Differential Equations means Ordinary Differential Equations.

falling off the Empire State Building, the initial distance or height may be $s_i = h$ and the final height may well be 0. But here, we start from $s = 0$ and integrate Eqn. 5-2.

$$\int_0^s \mathrm{d}s = \int_0^t v\mathrm{d}t = \int_0^t (v_i + at)\mathrm{d}t \quad \Rightarrow \quad \boxed{s = v_i t + \frac{1}{2}at^2} \qquad \text{Eqn. 5-3}$$

Problem: A cannonball is shot straight up from height, $z_i = 0$ with vertical velocity, $v_i = 110$ m·s^{-1}. Calculate 1: z at $t = 15$ s, and, 2: the time and value of its maximum height.
Information: When there is no friction, the only acceleration is gravity, $g \approx -10$ m·s^{-2}.
Solution: To find the height at any time, substitute $a = g$ and change s to z in Eqn. 5-3.

$$\boxed{z = v_i t + \frac{1}{2}gt^2} \qquad \text{Eqn. 5-4}$$

At $t = 15$ s,

$$z = 110(15) + \frac{1}{2}(-10)15^2 = 525 \text{ m}$$

The second part of the problem is more difficult because we must determine the time when the cannonball reaches its maximum height. There are several ways to solve this but in Calculus the word maximum should ring a bell instantly. The maximum height occurs when the derivative of height with respect to time, or vertical velocity, dz/dt = 0.

$$\frac{\mathrm{d}z}{\mathrm{d}t} = 0 = v_i + at_{\max} \rightarrow \quad t_{\max} = -\frac{v_i}{a} = -\frac{v_i}{g} = \frac{110}{10} = 11 \text{ s}$$

Now that we know $t_{\max}$, substitute into Eqn. 5-4 to find $z_{\max}$.

$$z_{\max} = v_i t_{\max} + \frac{1}{2}at_{\max}{}^2 = 110 \times 11 + \frac{1}{2}(-10)11^2 = 605 \text{ m}$$

Problem: The cannon is now tilted away from vertical and shot again with the same vertical velocity as in the previous problem, $v_i = 110$ m·s^{-1} but now with horizontal velocity, dx/dt = u_i = 55 m·s^{-1}. Assuming the cannon is located at $x = 0$, calculate the horizontal distance, x the cannonball has travelled when it hits the ground again (i. e., $z_f = z_i = 0$).
Information: In this case of no friction, the vertical equation of motion does not change when horizontal motion is added, so Eqn. 5-4 and its solutions remain valid.
Solution: When there is no friction, there is no horizontal force so the horizontal equation of motion has no acceleration. Substituting in Eqn. 5-3 with x for horizontal distance yields,

$$x = u_i t \qquad \text{Eqn. 5-5}$$

To find the final horizontal distance, x_f that the cannonball has travelled when it hits the ground, first set $z_f = 0$ in Eqn. 5-4 to find the final time, t_f. Then substitute into Eqn. 5-5.

$$z_f = 0 = v_i t_f + \frac{1}{2}gt_f{}^2 \quad \Rightarrow \quad t_f = -\frac{2v_i}{g} = 22 \text{ s} \quad \Rightarrow \quad x_f = 1210 \text{ m}$$

Figure 5-1 shows the trajectory of the cannonball including its location every second. The top frame is the solution with no friction or air resistance, and the bottom frame includes air resistance.

When there is no friction, the trajectory is a parabola and we prove it by substituting $t = x/u_i$ from Eqn. 5-5 into Eqn. 5-4.

$$z = \frac{v_i x}{u_i} + \frac{1}{2} g \left(\frac{x}{u_i} \right)^2$$

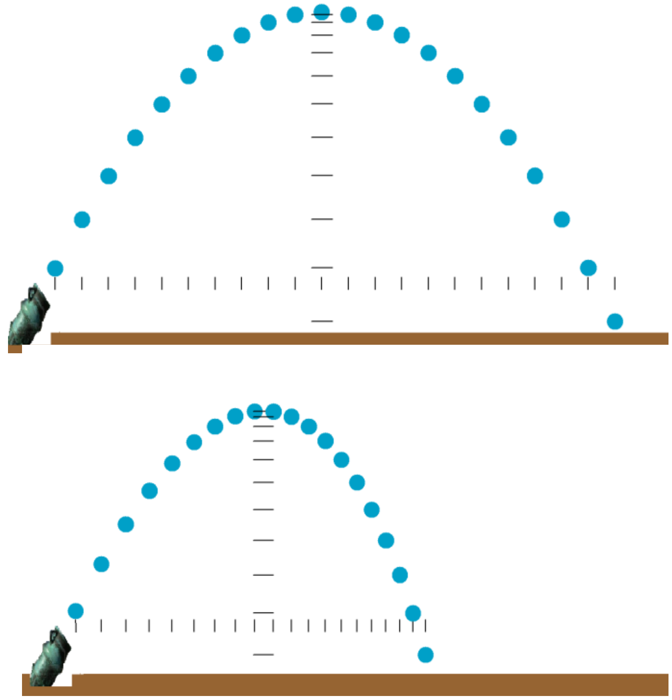


Figure 5-1 Trajectory of a cannonball without (top) and with (bottom) air resistance. Hatch marks and balls are shown at 1 s intervals.

The vertical and horizontal hatched lines are spaced at 1 s intervals in Figure 5-1. The distance between adjacent vertical lines indicates horizontal velocity. When there is no friction the lines are equally spaced because u is constant. The distance between adjacent horizontal lines indicates vertical velocity and shows that it is largest near the ground and zero at the top of the trajectory. Measure carefully and you will see that the vertical distance between adjacent lines increases by the same constant value from top to bottom. This is a consequence of constant acceleration.

With air resistance, the trajectory is not symmetric about its top point. Air resistance slows the vertical and horizontal components of the velocity so that the cannonball does not get as far or as high, hits the ground sooner (after about 20 s), and travels a shorter distance on the downward part of its trajectory. (This case was solved numerically.)

Problem: How fast will King Kong be falling just before he hits the ground after falling off the Empire State Building?

Information: The Empire State Building without the TV Tower is 380 m high.

Solution: This is a tougher problem because we must first solve Eqn. 5-4 for time. This gives a quadratic equation. Then, using t, we must solve Eqn. 5-2 for v.

Of course, we could do all this work – it is ‘just’ algebra – but let’s take a longer route that will make it simpler in the end because it gives a general result. For that we must return to the definitions of velocity and acceleration, and use the Chain Rule of Calculus.

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t} = v\frac{\mathrm{d}v}{\mathrm{d}s} = \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{v^2}{2}\right)$$

Multiplying by ds and integrating both sides of the equation yields the equation that relates distance and velocity.

$$\int_0^s a\,\mathrm{d}s = \int_{v_i}^{v_f} \mathrm{d}\left[\frac{v^2}{2}\right] \quad \Rightarrow \quad \boxed{{v_f}^2 = {v_i}^2 + 2as} \qquad \text{Eqn. 5-6}$$

Any object begins falling with $v_i = 0$, so the solution for v_f is,

$$v_f = \pm\sqrt{2gs} \approx -87\ \mathrm{m \cdot s^{-1}}$$

When King Kong hit the ground he would be falling at 87 $\mathrm{m{\cdot}s^{-1}} \approx$ 180 mph neglecting friction (which would not have slowed him much because he was so large) so had he fallen he would have died even if he hadn't been shot. Ironically, if an ant had fallen off the Empire State Building at the same time, it would have fallen much slower and taken much longer because of friction and would have walked away unharmed. This brings us to...

5.2 Motion Case #2: Motion Affected by Friction Only

Friction acts to reduce all relative motions. Its details are unbelievably complicated so physicists have created several simplifications. The most common is to assume that the friction force and acceleration are proportional to the velocity, and act to decelerate motion. This leads directly to the differential equation,

$$\boxed{\frac{\mathrm{d}v}{\mathrm{d}t} = -k_{fr}v} \qquad \text{Eqn. 5-7}$$

Eqn. 5-7 is a differential equation with a constant coefficient - in this case, the friction coefficient, k_{fr}. It is an equation that can be rearranged and integrated as follows;

$$\frac{\mathrm{d}v}{v} = -k_{fr}\mathrm{d}t \quad \Rightarrow \quad \int_{v_i}^{v} \mathrm{d}[\ln(v)] = \ln\left(\frac{v}{v_i}\right) = -k_{fr}t$$

Remember that the exponential is the inverse of the natural log. Therefore, v is given by

$$\boxed{v = v_i e^{-k_{fr}t}} \qquad \text{Eqn. 5-8}$$

Thus, v decreases exponentially with time when friction is proportional to v.

Exponential decay, introduced in Section 2.3 and described by Eqns. 5-7 and 5-8, applies to a vast range of phenomena including 1: radioactive decay, 2: chemical reaction rates that depend on the concentration of one chemical, 3: the decrease of pressure with height in the atmosphere (when temperature is constant), 4: decrease of the temperature difference with time between two points in a solid due to conduction, 5: the decrease of sunlight intensity with distance as it passes through a uniform fluid or fog, 6: the decrease of the real value of money you keep in your pillow with time when there is a constant rate of inflation and, 7: the electric charging rate of a capacitor with time. The first electrons enter the capacitor easily, but repel subsequent electrons. As a result, the capacitor charges more slowly as it approaches a fully charged state. For a similar reason, this probably also describes how your rate of eating slows as your stomach fills.

Differential equations with constant coefficients (such as Eqn. 5-5) always have solutions that either oscillate or decay or grow exponentially or both. This makes it worth taking a short diversion to consider the…

5.3 Cook Book Method for Solving Differential Equations with Constant Coefficients

Students quickly learn the simple, universal cook book technique for solving differential equations with constant coefficients. The underlying principle is that exponentials are the functions that are proportional to their derivatives. The resulting technique is to assume that $y(x) \propto e^{kx}$ ($\propto$ means proportional to). When this is done, we replace the 1st derivative by $k^1 y(x)$, the 2nd derivative by $k^2 y(x)$, and so on. The result is that the differential equation is replaced by an algebraic equation for the constant, k, and algebra equations are usually much easier to solve than differential equations.

The most general second order differential equation with constant coefficients is,

$$\boxed{\frac{d^2 y}{dx^2} + C\frac{dy}{dx} + Dy = f(x, y)} \quad \text{Eqn. 5-9}$$

The function, f(x, y) makes a differential equation nonhomogeneous. Eqn. 5-9 can be almost impossible to solve exactly if f(x, y) is a complicated function and is still tricky to solve even if f(x, y) is as simple as a constant. So, let's start real simple by considering the case where f(x, y) = 0. Then Eqn. 5-9 simplifies to the homogeneous form,

$$\boxed{\frac{d^2 y}{dx^2} + C\frac{dy}{dx} + Dy = 0} \quad \text{Eqn. 5-10}$$

Now we can proceed. Setting $y \propto e^{kx}$ converts the differential equation, Eqn. 5-10 into a quadratic algebraic equation for the unknown constant, k.

$$\left[k^2 + Ck + D\right]e^{kx} = 0 \quad \text{Eqn. 5-11}$$

Since e^{kx} is never zero, the quadratic must always be zero. So, we solve the quadratic equation for k just as we would in elementary algebra. This yields the two solutions,

$$k_1 = \frac{-C + \sqrt{C^2 - 4D}}{2} \qquad k_2 = \frac{-C - \sqrt{C^2 - 4D}}{2} \quad \text{Eqn. 5-12}$$

Because there are two distinct solutions for k whenever the discriminant is not 0 ($C^2 \neq 4D$) the solution for $y(x)$ looks like,

$$\boxed{y = Ae^{k_1 x} + Be^{k_2 x}} \quad \text{Eqn. 5-13}$$

The two constants, A and B are constants of integration that arise because the differential equation includes a second order derivative, which must be integrated twice. Case #4 and Case #5 will give examples of how to find A and B. I bet you can't wait!

5.4 Motion Case #3: Motions Affected by Gravity and Friction

Don't forget the tiny ant that fell off the Empire State Building with King Kong. Of course we didn't see it! Nor did we see it walk away unharmed. An ant is so small that air resistance does not allow it to fall fast. Furthermore, when an ant falls, it spreads itself out to slow its fall even more by producing extra resistance. As a result, a small to medium size ant reaches a maximum fall speed (called Terminal Velocity for several obvious reasons) of ≈ 2 m·s^{-1} ≈ 4 mph. This is slower than the terminal velocity of a raindrop (about 5-8 m·s^{-1}), another reason the ant doesn't splatter on hitting the ground.

All objects that are dropped and then fall through the atmosphere or through a liquid accelerate at first. Then, as the velocity increases, so does the retarding force of friction. Ultimately, there is no net force on the object and by Newton's Law of Motion, the object stops accelerating and reaches its terminal velocity.

Now that you know the behavior of a falling object that is slowed by friction, let's set up and solve its differential equation of motion, assuming that the friction force is proportional to the velocity. Acceleration is due to gravity (remember that g ≈ -10 m·s^{-2}) and friction, so,

$$\boxed{\frac{\mathrm{d}v}{\mathrm{d}t} = -k_{fr}v + g} \qquad \text{Eqn. 5-14a}$$

The constant, g makes Eqn. 5-14a nonhomogeneous and surprisingly tricky to solve. Rearranging to isolate the nonhomogeneous term on the right hand side is the first step toward solving the equation.

$$\boxed{\frac{\mathrm{d}v}{\mathrm{d}t} + k_{fr}v = +g} \qquad \text{Eqn. 5-14b}$$

While you are struggling to think of how to solve this new equation I will give you a hint by telling you the Aesop's Fable of,

The Fox and the Cat

A Fox was boasting to a Cat of its clever devices for escaping its enemies. "I must have a hundred ways of escaping my enemies. How many do you have?" "I have only one," said the Cat. Just then they heard the cry of an approaching pack of hounds. The Cat immediately scampered up a tree and hid in the boughs. "This is my way," said the Cat. "Which way are you going to use?" The Fox thought first of one way, then of another, and while he was debating the hounds came nearer, and at last the Fox in his confusion was caught by the hounds and torn to pieces.

The moral: **"Better one sure way than a hundred on which you cannot reckon."**

Does that help? It should! Let's try our old but reliable trick of transforming the left hand side of Eqn. 5-14b into a perfect differential so that it can be integrated. And, if the right hand side is simple enough, as it is Eqn. 5-14b, it can also be integrated.

The left hand side would be a perfect differential if it were equal to

$$u\frac{\mathrm{d}v}{\mathrm{d}t} + v\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\mathrm{d}(uv)}{\mathrm{d}t}$$

A function, u that converts a differential equation into a form that can be integrated is called an **Integrating Factor**.

The procedure to convert a differential equation like Eqn. 5-14b into a form that can be integrated involves two steps, 1: multiply each term in the equation by a function, u, that we have not yet determined and, 2: insist that the second term in the equation is equal to $v \cdot \mathrm{d}u/\mathrm{d}t$. In that case Eqn. 5-14b becomes,

$$u\frac{\mathrm{d}v}{\mathrm{d}t} + vk_{fr}u = u\frac{\mathrm{d}v}{\mathrm{d}t} + v\frac{\mathrm{d}u}{\mathrm{d}t} = gu \qquad \text{Eqn. 5-15}$$

Eqn. 5-15 is true only if Eqn. 5-16 (the differential equation for the unknown, u) is also true,

$$k_{fr}u = \frac{\mathrm{d}u}{\mathrm{d}t} \qquad \text{Eqn. 5-16}$$

Because Eqn. 5-16 has constant coefficients, its solution, u, is exponential, namely

$$u = Ae^{k_{fr}t} \qquad \text{Eqn. 5-17}$$

Substituting this value of u into Eqn. 5-15 gives it the form,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[e^{k_{fr}t}v\right] = ge^{k_{fr}t} \qquad \text{Eqn. 5-18}$$

Now, simply, yes simply, multiply each side of Eqn. 5-18 by dt and integrate the left integral from 0 to v and the right integral from 0 to t.

$$\int_0^v \mathrm{d}\left[e^{k_{fr}t}v\right] = \int_0^t ge^{k_{fr}t}\mathrm{d}t \quad \Rightarrow \quad e^{k_{fr}t}v - 0 = \frac{g}{k_{fr}}\left(e^{k_{fr}t} - 1\right)$$

The rest is JUST algebra. Simply (again) solve for v.

$$\boxed{v = \frac{g}{k_{fr}}\left(1 - e^{-k_{fr}t}\right)} \qquad \text{Eqn. 5-19}$$

Note that at t = 0, v = 0 and that as $t \to \infty$, v approaches terminal velocity,

$$v(t \to \infty) \to \frac{g}{k_{fr}} < 0$$

Figure 5-2 shows the graph of v from Eqn. 5-19. It looks exactly like exponential decay except that v asymptotically approaches a terminal velocity, which is negative because gravity ($g \approx$ -10 m·s^{-2}) is downward.

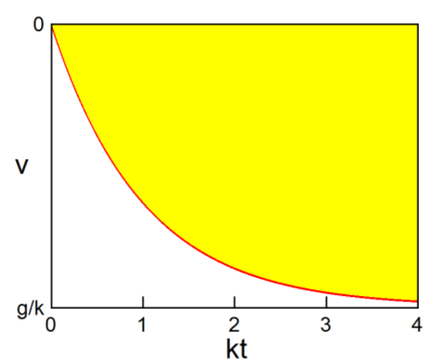


Figure 5-2 Fall speed when friction is proportional to v.

5.5 Motion Case #4: Motion of a Mass on a Spring: NO Friction

Spring motion represents a good introduction to waves because when a spring is stretched from its relaxed or neutral position and released, it will pull back. If there is no friction, the spring will not stop at the neutral position but will overshoot it and keep on oscillating forever. Swings behave similarly. The neutral position for a swing is when the chain is vertical and the seat is at the lowest position. If you push the swing forward it will be forced back and will not stop at the lowest point but will continue backward. If there is no friction the swing will continue backward until it is exactly as far back as it had been forward. Small children pushing their friends or younger siblings for the first time utterly fail to anticipate the backward motion and get knocked over, giving great joy to all sadists.

This experience leads us to conclude that there is a force that is directed toward the neutral point. For springs this restoring force is closely proportional to the distance, *x* from the neutral point. This relation was discovered by Robert Hooke in 1660 and is called Hooke's Law. It applies so long as the spring is not stretched too far or snapped. Hooke's Law then leads to the spring equation,

$$F = ma = \boxed{m\frac{\mathrm{d}v}{\mathrm{d}t} = -k_{sp}x} \qquad \text{Eqn. 5-20}$$

Why is nothing easy in Physics? It seems we have one equation, but two unknowns, *v* and *x*. We are in luck if we remember that $v = \mathrm{d}x/\mathrm{d}t$, because then Eqn. 5-20 transforms to,

$$\boxed{m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -k_{sp}x} \qquad \text{Eqn. 5-21}$$

If you stare at Eqn. 5-21 for a few moments without panicking, you might notice **1**: that *x* is proportional to minus its second derivative. If you made this crucial observation you might also remember **2**: that sines and cosines are the functions that are proportional to minus their second derivatives. This would represent great insight on your part and if you made the connection you should be quite proud of yourself. In any case, Eqn. 5-21 has constant coefficients, so $x \propto e^{kt}$, and standard procedure is to substitute e^{kt} for *x* to get the quadratic equation, Eqn. 5-22 for the unknown, *k* (which is not the same as k_{sp}).

$$\left(k^2m + k_{sp}\right)e^{kt} = 0 \qquad \text{Eqn. 5-22}$$

Since e^{kt} is never 0, the terms in parenthesis are. Then, solving for *k* yields,

$$k = \pm\sqrt{-\frac{k_{sp}}{m}} = \pm i\sqrt{\frac{k_{sp}}{m}} \qquad \text{Eqn. 5-23}$$

Oh, my! The discriminant is negative! That makes *k* imaginary. But all is not lost. Thank Goodness we remember from Section 4.2 that exponentials raised to imaginary powers are sines and/or cosines. So, the formal procedure started with exponentials, but our insight showed that it would ultimately be some combination of sines and cosines.

This means the solution consists of waves. And, of course, that is exactly what the back and forth motion of springs and swings means. So, the solution to Eqn. 5-21 is,

$$x(t) = C_1 e^{i\sqrt{\frac{k_{sp}}{m}}t} + C_2 e^{-i\sqrt{\frac{k_{sp}}{m}}t} = A\cos\left(\sqrt{\frac{k_{sp}}{m}}t\right) + B\sin\left(\sqrt{\frac{k_{sp}}{m}}t\right) \qquad \text{Eqn. 5-24}$$

Problem: This spring problem has 4 parts. A spring fixed to a ceiling stretches 0.05 m when a 0.1 kg mass is attached to it. **1**: Find the spring constant. **2**: When the spring is stretched an additional 0.1 m and then released from a state of rest, solve for *x* and *v* as functions of time. **3**: Find the period and frequency of the motion. **4**: Find the period and frequency when the attached mass is increased by a factor of 10 to 1 kg.

Solution: **Step 1**: When the mass is attached to a hanging spring it causes a downward force equal to its weight, *mg*. When the situation is balanced, the spring will stretch until the spring force offsets the weight. This occurs when $x = -0.05$,

$$mg = k_{sp}x \quad \Rightarrow \quad k_{sp} = \frac{mg}{x} = \frac{0.1(-10)}{-0.05} = 20$$

Step 2: At $t = 0$, the spring is stretched an additional $x = 0.1$ m and released from rest so, $v(0) = 0$. Substituting all values into Eqn. 5-24 at $t = 0$ yields the value of *A*.

$$x(t=0) = A\cos(0) + B\sin(0) = 0.1 \quad \Rightarrow \quad A = 0.1$$

Unfortunately, this equation does not give any information about the constant *B*. Of course, we can only determine one constant since we have only used one fact – the initial position is 0.1 m. When we use the 2nd fact that the initial velocity, $v(0) = 0$ we will be able to find *B*. How to do this? Maybe you remember that $v = dx/dt$. So, take the derivative of Eqn. 5-24 and set it equal to 0 when $t = 0$. This gives $B = 0$. (You do it.) Therefore,

$$x(t) = 0.1\cos\left(\sqrt{\frac{k_{sp}}{m}}t\right) \qquad \text{Eqn. 5-25}$$

Step 3: The wave period equals the time it takes to increase the angle by 2π since this defines one cycle. In other words, $\cos(\theta + 2\pi) = \cos(\theta)$. Applying this to the angle in Eqn. 5-25 and using τ as the time for 1 cycle, yields,

$$2\pi = \sqrt{\frac{k_{sp}}{m}}\tau \quad \Rightarrow \quad \boxed{\tau = 2\pi\sqrt{\frac{m}{k_{sp}}}} \qquad \text{Eqn. 5-26}$$

For $m = 0.1$ kg and $k_{sp} = 20$, the period, $\tau_{0.1\text{ kg}} = 0.444$ s. Frequency is the inverse of period, so there are $1/0.444 \approx 2.25$ cycles per second.

Step 4: Increasing *m* also increases period, τ (because springs, unlike gravity, accelerate more massive objects more slowly.) Eqn. 5-26 shows that $\tau \propto m^{1/2}$, so, increasing *m* by a factor of 10 increases τ by a factor $\sqrt{10} \approx 3.16$. Thus $\tau_{1\text{ kg}} \approx 3.16 \times \tau_{0.1\text{ kg}} \approx 1.40$ s.

5.6 Motion Case #5: Motion of a Mass on a Spring: YES Friction

Friction will not only slow spring motion but may even change its nature. If you give a swing a single push, it will oscillate back and forth with decreasing amplitude and eventually come to rest. If the swing were immersed in oil instead of air, it might never reach the neutral position, but ooze toward it. To get the governing differential equation for spring motion with friction, simply add the friction force to the right hand side of Eqn. 5-21

$$\boxed{m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -k_{sp}x - mk_{fr}\frac{\mathrm{d}x}{\mathrm{d}t}} \qquad \text{Eqn. 5-27a}$$

Rearranging

$$\boxed{\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + k_{fr}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{k_{sp}}{m}x = 0} \qquad \text{Eqn. 5-27b}$$

Assuming exponential solutions for x leads to the quadratic algebra equation,

$$k^2 + k_{fr}k + \frac{k_{sp}}{m} = 0$$

The solution of the quadratic is,

$$k = -\frac{k_{fr}}{2} \pm \sqrt{\frac{k_{fr}^2}{4} - \frac{k_{sp}}{m}} \qquad \text{Eqn. 5-28}$$

With friction there are two distinct types of motion depending on whether the discriminant in Eqn. 5-28 is positive or negative. If it is positive, the motion is pure exponential decay, as in Eqn. 5-29.

$$\boxed{x(t) = 0.1e^{-\left[\frac{k_{fr}}{2} \pm \sqrt{\left(\frac{k_{fr}^2}{4} - \frac{k_{sp}}{m}\right)}\right]t}} \qquad \text{Eqn. 5-29}$$

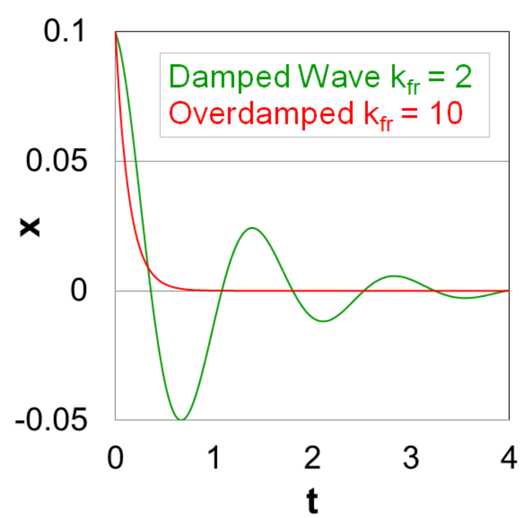


Figure 5-3 Damped waves (green) and overdamped motion (red).

Exponential decay, called **overdamped** motion, occurs when the friction constant is large enough. The motion is shown by the red curve in Figure 5-3 when k_{fr} = 10. The spring rapidly relaxes back to the neutral point, x = 0 but does not overshoot it.

Alert 1: Eqn. 5-29 has two possible solutions. The solution with a minus sign in front of the radical in the brackets is not physically realistic because as k_{fr} increases it damps out the motion more slowly. Therefore, it is discarded.

If the friction constant is small enough so that the discriminant is negative, the motion consists of a damped cosine wave given by Eqn. 5-30 that decays exponentially with time.

When k_{fr} = 2 the motion is shown by the green curve in Figure 5-3.

$$\boxed{x(t) = 0.1e^{-\frac{k_{fr}}{2}t}\cos\left(\sqrt{\left(\frac{k_{sp}}{m} - \frac{k_{fr}^2}{4}\right)}t\right)} \qquad \text{Eqn. 5-30}$$

Alert 2: The terms in the discriminant of Eqn. 5-30 appear in reverse order from Eqn. 5-28 and 5-29.

An intriguing analogy to spring motion with friction is the strange phenomenon of tunneling in Quantum Mechanics described by Schrödinger's Equation. When electrons encounter energy barriers that would 'normally' be too high to jump over they have a probability (less than 100%) of penetrating or tunneling through these barriers that resembles overdamped motion described by Eqn. 5-29. On either side of the barrier the electrons behave like waves with no friction. But with this tantalizing hint of what you can learn after mastering Calculus we have tunneled far enough.

5.7 Flooding and Overdosing

Flooding rivers act much like overdosing bodies (or spreading epidemics). Rivers often crest long after the rain has ended or subsided. In short streams, peak floods occur hours after the rain. For progressively longer rivers, peak floods occur days, weeks, or up to three months after the rainy season for extremely long rivers such as the Nile. It takes time for rain to soak into the ground or run along the ground and reach the rivers and then more time for the water to flow hundreds of miles downstream.

A similar lag occurs when we eat or take drugs. When we wolf down food or take drugs (by mouth), it doesn't register immediately that we have overeaten or overdosed. It takes the digestive system roughly an hour to process food or drugs and fill the blood with sugar and other chemicals. But since it takes even longer for the cells to remove the sugar and other chemicals (e. g., half-life, $t_{½} \approx 5$ h for caffeine in the blood system), concentrations may increase in the blood system for about an hour before beginning to decrease. I cannot tell you how many times I have eaten chocolate shortly before going to sleep at night and then woke sleeplessly and regretfully from the caffeine kick an hour or two later. Then it would typically take several hours to get back to sleep.

The differential equation that describes rapid flooding or rapid input to any system with a slower output or processing time is a form of the Fundamental Equation of Systems, Eqn. 1-33. For a reservoir or any section of a stream, the rate of change of volume, V of the water equals the volume inflow rate, Q_{in} minus the volume outflow rate, Q_{out}.

Rate of Change = Inflow Rate – Outflow Rate

$$\boxed{\frac{\mathrm{d}V}{\mathrm{d}t} = Q_{in} - Q_{out}} \qquad \text{Eqn. 5-31}$$

Volume, V equals surface area, A times height, h or **stage** of the stream or reservoir, i. e., $V = Ah$. If we assume that A is constant then $\mathrm{d}V/\mathrm{d}t = A\mathrm{d}h/\mathrm{d}t$. A is constant for a swimming pool but it increases with h in streams and reservoirs because banks slope outward. Neglecting that 'tiny' discrepancy from reality in the name of simplicity, Eqn. 5-31 becomes,

$$\boxed{\frac{\mathrm{d}h}{\mathrm{d}t}=\frac{Q_{in}-Q_{out}}{A}} \qquad \text{Eqn. 5-32}$$

Consider that the inflow comes from a pulse of rainfall plus an almost constant low level background flow called base flow, Q_{base}. Rain intensity is so irregular with alternating intermittent downpours and lulls that it is difficult to give it any neat function. I simplify it by assuming that it is a large pulse that decreases rapidly and exponentially with time. Thus,

$$Q_{in}=R_0e^{-k_Rt}+Q_{base}$$

The outflow rate increases as h increases and is proportional to some power of h.

$$Q_{out}=K_sh^p$$

A typical value is $p \approx 1.5$. Surely, $p > 1$ since Q_{out} more than doubles when h doubles because the current is faster and the river is also wider. With these expressions for Q_{in} and Q_{out}, Eqn. 5-32 becomes,

$$\boxed{\frac{\mathrm{d}h}{\mathrm{d}t}=\frac{R_0e^{-k_Rt}+Q_{base}-K_sh^p}{A}} \qquad \text{Eqn. 5-33}$$

This differential equation is too difficult for us to solve analytically unless we simplify by setting $p = 1$ and $Q_{base} = 0$. In that simple case, Eqn. 5-33 reduces to,

$$\boxed{\frac{\mathrm{d}h}{\mathrm{d}t}+\frac{K_s}{A}h=\frac{R_0e^{-k_Rt}}{A}} \qquad \text{Eqn. 5-34}$$

Do you remember how to solve a nonhomogeneous differential equation with constant coefficients? Remember Case #3: Falling Motion with Friction? Remember the Fable of the Fox and the Cat? Remember Eqn. 5-14b, and the Integrating Factor? So, let's proceed.

Step #1: Multiply each term in Eqn. 5-34 by the unknown integrating factor, u, (which is a function of t).

$$u\frac{\mathrm{d}h}{\mathrm{d}t}+h\frac{K_s}{A}u=\frac{R_0e^{-k_Rt}}{A}u \qquad \text{Eqn. 5-35}$$

Step #2: Then $u(t)$ must satisfy the differential equation,

$$\frac{\mathrm{d}u}{\mathrm{d}t}=\frac{K_s}{A}u$$

Step #3: Solve the differential equation for $u(t)$.

$$u = u_0 e^{\frac{K_s}{A}t} \qquad \text{Eqn. 5-36}$$

Step #4: Substitute u from Eqn. 5-36 into Eqn. 5-35. (Note that u_0 cancels.)

$$\int_{h_0}^{h} \mathrm{d}\left[h e^{\frac{K_s}{A}t} \right] = \int_0^t \frac{R_0 e^{-k_R t}}{A} e^{\frac{K_s}{A}t} \mathrm{d}t = \frac{R_0}{A} \int_0^t \mathrm{e}^{-\left(k_R - \frac{K_s}{A}\right)t} \mathrm{d}t \qquad \text{Eqn. 5-37}$$

Step #5: Integrate both sides of Eqn. 5-37 and rearrange to solve for $h(t)$.

$$h = e^{-\frac{K_s}{A}t} \left[h_0 + \frac{R_0}{A\left(k_R - \frac{K_s}{A}\right)} \left[1 - e^{-\left(k_R - \frac{K_s}{A}\right)t} \right] \right] \qquad \text{Eqn. 5-38}$$

Eqn. 5-38 is the exact solution of Eqn. 5-34. (Eqn. 5-38 may look evil but just has a few exponential terms.) The constants, $h_0 = 0$, $R_0 = 1000$, $k_r = 0.05$ $K_s = 10$, and, $A = 1000$ are not arbitrary. In order to describe a system that has a rapid input and slow response or output, it is necessary that $k_r >> K_s/A$. I did not specify the units so that time could be anywhere from seconds to months, depending on the situation, which a hydrologist or physician would know.

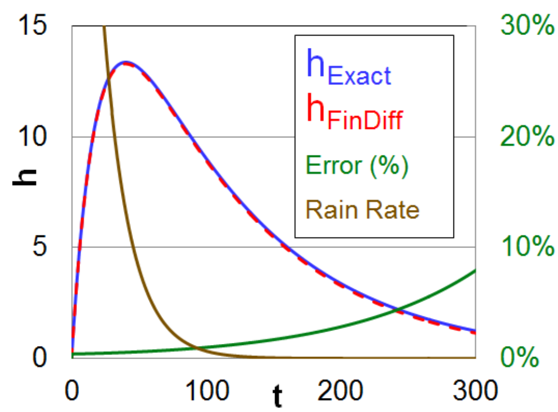


Figure 5-4 Exact (blue) and finite difference (red) solutions for h, % error (green) and rain rate (brown).

Figure 5-4 displays the time evolution of Eqn. 5-38. It is a busy graph, so focus on one curve at a time. The blue curve is the exact solution for water height. It rises to a maximum in a relatively short time and then gradually settles back toward the original value. This curve tells us both the time and height of the maximum flood so that the public can be warned and protective measures can be taken to minimize damage. Thus, Eqn. 5-38 is an invaluable prediction equation for floods. The brown curve represents the rainfall rate, which is so large at first that it doesn't fit in the graph, but rapidly decreases, effectively ending before the flood crests, as is often the case for long rivers. The red dashed curve represents the finite difference solution that we will do in Section 5.9. The green curve (which uses the scale on the right) represents the percent error of the finite difference solution. Errors of finite difference solutions often grow with time, as this one does. Errors of weather forecasts made by finite difference equations always grow with time until they lose all accuracy and skill. Errors of economic forecasts made with any equations do not grow with time simply because they have no skill or accuracy to begin with. (That's a nasty comment but you decide if it is true!)

5.8 Population Growth: The Logistic Differential Equation

If there is no limit to animal (or plant) populations, the animals reproduce in proportion to their population, P at a rate $R \times P$, where R is the reproduction rate. The population therefore obeys Eqn. 2-10 and grows exponentially following Eqn. 2-11. In reality, near exponential growth may continue for a while, but sooner or later limits to population crop up. The food supply begins to run out and predators or parasites proliferate. One simple way to include these limiting factors is to assume that there is a limiting stable population, P_{max} and that population growth rate is proportional to the difference between the limiting stable population and the actual population, or (P_{max} - P). With this additional factor (and the change of letters), Eqn. 2-10 is modified to become the (now) famous

Logistic Differential Equation

$$\boxed{\frac{\mathrm{d}P}{\mathrm{d}t} = RP(P_{\max} - P)} \qquad \text{Eqn. 5-39}$$

This we solve with the cookbook technique of separating variables and integrating.

$$\int_{P_0}^{P} \frac{\mathrm{d}P}{P(P_{\max} - P)} = \int_0^t R\mathrm{d}t = Rt \qquad \text{Eqn. 5-40}$$

The left hand integral of Eqn. 5-40 gives us only minor troubles if we remember to find partial fractions of the integrand.

$$\frac{1}{P(P_{\max} - P)} = \frac{A}{P} + \frac{B}{(P_{\max} - P)} = \frac{A(P_{\max} - P) + BP}{P(P_{\max} - P)}$$

Equating the numerators of the extreme left and extreme right hand sides produces two equations for A and B,

$$1 = A(P_{\max} - P) + BP \quad \Leftrightarrow \quad \begin{array}{l} 1 = AP_{\max} \\ 0 = B - A \end{array} \quad \Leftrightarrow \quad A = B = \frac{1}{P_{\max}}$$

Then, the left hand side of Eqn. 5-40 integrates to yield,

$$\int_{P_0}^{P} \frac{\mathrm{d}P}{P(P_{\max} - P)} = \frac{1}{P_{\max}} \left[\int_{P_0}^{P} \frac{\mathrm{d}P}{P} + \int_{P_0}^{P} \frac{\mathrm{d}P}{(P_{\max} - P)} \right] = \frac{1}{P_{\max}} \ln\left[\frac{P(P_{\max} - P_0)}{P_0(P_{\max} - P)} \right] = Rt$$

Taking antilogs (exponentials) and solving this equation for P (You do the algebra!) yields,

$$\boxed{P = \frac{P_{\max}}{1 + \left(\frac{P_{\max}}{P_0} - 1 \right) e^{-RP_{\max} t}}} \qquad \text{Eqn. 5-41}$$

The solution for P from Eqn. 5-41 is graphed in Figure 5-5 for the case of $P_{max} = 1$, $P_0 = 0.002$ and, $R = 3.8$. The curve rises exponentially at first. Its slope is greatest at $P_{max}/2 = 0.5$ and it levels off (exponentially) as P approaches its maximum value, P_{max}. Did you realize that P is a hyperbolic tangent?

Warning! I hope that Figure 5-5 makes you feel that, as Hamlet put it, "Something is rotten in the state of Denmark!" Let's put this on trial because we have several objections. The only doubt about the following trial is the assumption that Judges or Attorneys have any knowledge or understanding of science or math.

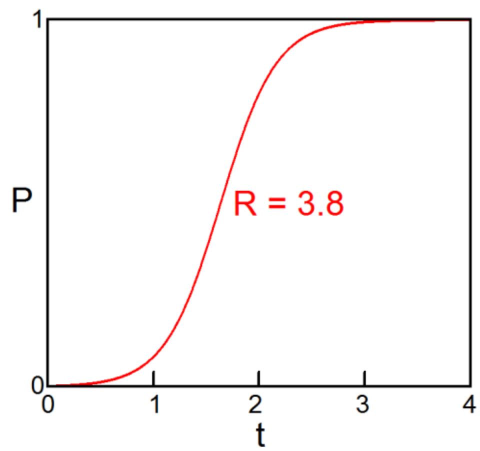


Figure 5-5 Population density for the Logistic Diff. Eq. when $P_{max} = 1$, $P_0 = 0.002$, $R = 3.8$.

Attorney: "Objection, Your Honor! Figure 5-5 and indeed the entire argument cannot be right because you can't have a fraction of an animal.

Judge: "Objection overruled! It is true that you can't have 0.5 animals, but P in Eqn. 5-41 represents *population density*. Then there is no objection to its having fractional values. For example, a horse requires about 3 acres of grass so that $P_{max} \approx ⅓$ horse per acre.

Attorney: "Objection, Your Honor!"

Judge: "What, another objection?"

Attorney: "Yes! And this one is far more serious! No animal can have continuous births. No creatures can give birth until they reach maturity, which takes time, and when they give birth, they may have many offspring at one time. Birth, like earning interest is episodic. Animals do not compound continuously. In fact, many animals give birth only once a year and have a huge number of offspring.

Judge: "Objection sustained!" But I don't expect this will change the solution very much!

Attorney: "Objection, Your Honor!"

Judge: "What? Another objection! You are getting to be a real pain in the... um, neck!"

Attorney: "Why not? After all, I am a lawyer! This objection is extremely serious. The solution Eqn. 5-41 to the differential equation Eqn. 5-39 is always a beautiful, smooth hyperbolic tangent. But real animal populations sometimes experience boom and bust cycles, which may be erratic. So there is something fundamentally wrong with Eqn. 5-39 after all!

Judge: "Objection partly sustained and partly overruled! And, since we have a finite difference of opinion, you must create a finite difference version of Eqn. 5-39 and solve it by the finite difference technique of iteration.

Attorney: "I am a lawyer, not a mathematician. I can't create or solve anything. I'll have to hire some math nerds who can, and of course underpay them."

5.9 Return to Finite Differences: The Logistic Difference Equation and Chaos

The preceding two sections show that there are at least two reasons to return to finite differences. **First**, some functions and differential equations are so complicated that they are impossible to solve analytically. Finite difference techniques may not be perfectly accurate, but they work. **Second**, the attorney was right; some processes are not continuous but episodic, such as reproduction or compounding. In those cases the situation is governed by a finite difference equation and the continuous differential equation is not as accurate. In some cases even the fundamental nature of the solutions of the differential and finite difference equations are different!

Regarding the first reason (Nature's irregularity), just try walking across a stream and you will see how irregular the stream bed is! You may well stub your toes on a submerged rock or tire or fall in a hole like so many comedians. This example illustrates how mathematically complicated many of nature’s situations are. There are so many differential equations that cannot be solved exactly that when you find one you can solve, it is a cause for celebration. But it is also a grand opportunity to use that equation and its solution for all it is worth and test the accuracy of numerical solutions.

Any differential equation, no matter how difficult, can be solved numerically, which usually means approximately. How accurate are the finite difference solutions? The only way you can be reasonably sure is to compare the numerical solution to a known, exact analytic solution. And that is the meaning and purpose of including the red dashed curve in Figure 5-4. It is the finite difference solution to Eqn. 5-34 when the time step, $\Delta t = 0.5$. The green curve shows the percentage error of the finite difference solution. The error is small but grows with time and depends on the size of the time step, Δt. For most equations, the smaller Δt, the more accurate the finite difference solution. In this case, the errors for short forecast times are almost exactly proportional to the size of the time step.

So now, we use our knowledge that the errors of the finite difference solution remain small for a long time after the maximum flood in Figure 5-4 with a time step, $\Delta t = 0.5$ to solve a similar but more difficult version of Eqn. 5-33 that is too hard for us to solve exactly. For this version, set $p = 1.5$ and include a base flow in the river, $Q_b = 30$, and keep the time step, $\Delta t = 0.5$. Then, the Fundamental Equation of Change is given by the finite difference version of Eqn. 5-33, or,

$$h_{t+\Delta t} = h_t + \frac{1}{A}\left[R_0 e^{-k_R t} - K_s h_t^{1.5} + Q_{base}\right]\Delta t \qquad \text{Eqn. 5-42}$$

Let’s take the first few steps. Remember from Section 1.7 that the technique is called iteration and it involves alternately solving and updating repeatedly. But first, there is one preliminary step, namely to find the initial height of the river due only to the base flow, $Q_b = 30 = K_s h_0^{1.5}$. Solving for h, yields, $h_0 = [30/K_s]^{1/1.5} = 2.08$.

Now, we are ready. Starting from $t = 0$, we solve for the height, h at $t = \Delta t = 0.5$.

$$h_{\Delta t} = 2.08 + \frac{1}{1000}\left[1000\, e^{-0.05(0)} - 10(2.08)^{1.5} + 30\right] \times 0.5 = 2.57$$

Now update and solve again for height, h at time, $t = 2\Delta t = 1.0$.

$$h_{2\Delta t} = 2.57 + \frac{1}{1000}\left[1000\, e^{-0.05(2)} - 10(2.57)^{1.5} + 30\right] \times 0.5 = 3.03$$

We could do this for many more steps, but I think you get the point.

Now to the second objection! What happens when the process is episodic, i. e., finite difference in nature rather than continuous? The finite difference form of Eqn. 5-39 is the...

The Logistic Difference Equation

$$\boxed{P_{n+1} = RP_n(1-P_n)} \quad \text{Eqn. 5-43}$$

In Eqn. 5-43, P_{n+1} is the future value of P, and it depends on the present value, P_n. Details of the first 5 steps of the iteration of Eqn. 5-43 are shown in Table 5-1 where the colored numbers at the end of each line are repeated in the calculations for the next time step.

n	P
0	P_0 = 0.002
1	P_1 = R×P_0×(1-P_0)=2×0.002000×(1-0.002000)=0.003992
2	P_2 = R×P_1×(1-P_1)=2×0.003992×(1-0.003992)=0.007952
3	P_3 = R×P_2×(1-P_2)=2×0.007952×(1-0.007952)=0.015778
4	P_4 = R×P_3×(1-P_3)=2×0.015778×(1-0.015778)=0.031058
5	P_5 = R×P_4×(1-P_4)=2×0.031058×(1-0.031058)=0.060186

Table 5-1 Five successive calculations of Eqn. 5-43 when P_0 = 0.002 and R = 2.

Eqn. 5-43 is extremely deceptive. The calculations are simple but you will soon see that the nature of the solutions is incredibly complex. All solutions to the differen**tial** equation, Eqn. 5-39 tend continuously toward the same asymptotic limit. The finite differen**ce** equation also has a steady solution. It is given by Eqn. 5-44 and is found by setting P_{n+1} = P_n in Eqn. 5-43.

$$\boxed{P_{eq} = 1 - \frac{1}{R}} \quad \text{Eqn. 5-44}$$

There are two differences between the asymptotic solution, P_{max} of the differen**tial** equation, 5-39 and the steady solution, P_{eq} of the differen**ce** equation, 5-43. First, P_{max} is the same for all R while P_{eq} varies with R. Second, and more important, not only might the solution, P to Eqn. 5-43 never reach or even approach the steady solution, but if R is large enough, the steady solution, P_{eq} is so unstable that any deviation from it will cause P to deviate widely from it.

An astounding property of the logistic difference equation, Eqn. 5-43 is the changing nature of its solutions for different ranges of R, outlined in Table 5-2 and shown in Figure 5-6 for 25 steps or years. R < 1 dooms the animals to extinction. Only for 1 < R < 3, does population reach a finite steady state. For R > 3 the behavior is more complex. For 3 < R < 3.449, the population settles into a simple alternation between 1 good and 1 bad period. Compound alternations occur for 3.449 < R < 3.57. Finally, for most values of R > 3.57 the population varies

RANGE OF R	NATURE OF THE SOLUTION
0 → 1	Extinction
1 → 2	Steady State
2 → 3	Asymptotic with Dying Waves
3 → 3.449	Simple Waves
3.449 → 3.57	Compound Waves
3.57 → 4	**CHAOS** (mostly)

Table 5-2 Nature of Solutions for the Logistic Difference Equation for various ranges of R.

erratically. Any tiny deviation in the initial population causes the eventual population to be totally different. The result is that, given any error in the initial measurement, no matter how tiny, it is impossible to make accurate predictions more than a few steps ahead.

This is the nature of scientific **CHAOS**, which governs a host of phenomena such as the weather, populations of animals with high reproduction rates (such as snowshoe hares), stock market fluctuations, cardiac fibrillation, epileptic seizures, eye movements in schizophrenics, and even brain waves in 'normal' people. If we humans were totally predictable we could not be creative.

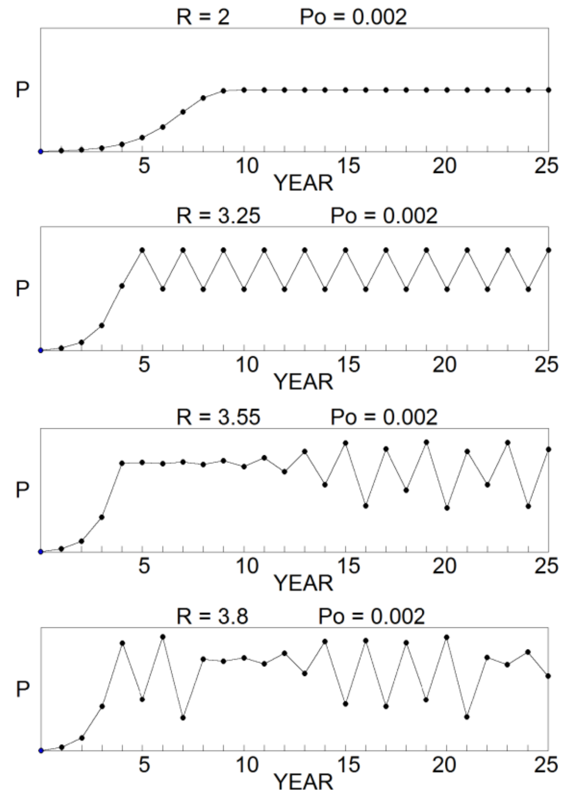


Figure 5-6 Solutions of the Logistic Difference Equation for R = 2, 3.25, 3.55 and 3.8.

A Subtle Limitation

Very quietly, the problems we treated using finite differences included differential equations with only first derivatives. But equations such as Eqn. 5-9, Eqn. 5-21 or Eqn. 5-27 include 2nd order derivatives. In that case, we need a finite difference expression for the 2nd derivative. Fortunately, we derived two such expressions back in Chapter 2, namely the forward difference form, Eqn. 2-9a and the centered difference form, Eqn. 2-9b. Solving higher order differential equations numerically requires more care, more ingenuity, more work, and more time, but the basic technique of iterating the Equation of Change still works. However, those equations remain the subject for a higher order book.

Summary

In this chapter we have learned that Ordinary Differential Equations (ODE's) involve a host of phenomena including motions, river levels, and animal populations. Because ODE's involve derivatives or rates of change they can be used to make predictions, and are usually solved by integration using a variety of techniques such as finding an integrating factor. When ODE's are too difficult to solve, they can be transformed into finite difference equations, which are solved numerically by iteration, and which may describe episodic phenomena such as animal populations more accurately.

The differential equations of this chapter are called Ordinary Differential Equations (ODE's) because they have only one independent variable. This has restricted us to a flat world. Now, it's on to higher dimensions and directions and vectors, more independent variables, and Partial Differential Equations (PDE's). Then, we'll take our chances about how the book will end!

CHAPTER 6: ADDING DIMENSIONS AND DIRECTION: MULTIVARIATE CALCULUS

6.1 Introduction to the 3-D World

So far (except in Section 3.11) we have worked in a world that is two dimensional – y is a function of x. The 2-D world is like a flat map with x extending to the east and y, to the north. It contains parabolas, hyperbolas, triangles, squares, polygons, circles and ellipses but not cubes, spheres, pyramids, cylinders, cones or helices. The 3-D world looks very different – it is far more beautiful and even awesome. It is a world with height and depth.

I remember the first time I saw the Grand Canyon. I drove north from Flagstaff, Arizona through the forested Coconino Plateau. For miles the road was almost level and trees blocked any distant view. Suddenly, the forest ended and the world dropped away into the depths of the Grand Canyon, as much as 5000 feet below. I gasped impulsively; I couldn't believe my eyes. I became one of the millions of witnesses to one of the most awesome, stupendous scenes on Earth, as in Figure 6-1.

Figure 6-1 At the edge of the Coconino Plateau and the Grand Canyon.

The 3-D world of math needs an extra independent variable. We already have x for east and y for north. We could add z for height, or some other variable, such as T for temperature or p for pressure. We could add a 4^{th} independent variable, such as time, t.

As soon as we more than one independent variable, we must introduce partial derivatives and partial differential equations. Keeping track of different directions makes it useful to phrase equations in terms of Vectors. These form the subjects covered in this chapter, namely Multivariate and Vector Calculus (plus the grand finale – Statistics).

Let's start climbing our way slowly through these advanced topics with a 3^{rd} variable, height, z, and consider…

6.2 Topographic Maps and the 3rd Dimension

Come and climb Double Sine Mountain. Figure 6-2 shows what it looks like both in rendered form (top) and as a topographic contour map (bottom). The mountain extends from $0 \le x \le \pi$ km and $0 \le y \le \pi$ km. Its height, z, depends on both x and y according to the equation,

$$z(x, y) = 1.05\sin(x)\sin(y) \qquad \text{Eqn. 6-1}$$

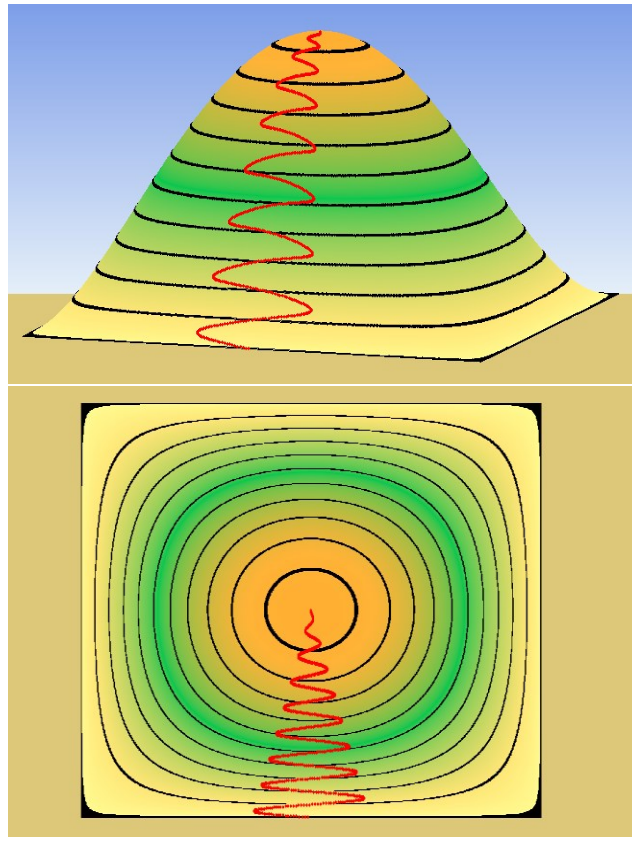


Figure 6-2 Rendered topographic map (top) and contour map (bottom) of Double Sine Mountain. Black contour lines are 100 m apart. The wavy red line is the snake path.

The minimum elevation, $z = 0$ occurs along the boundaries where both $\sin(0) = 0$ and $\sin(\pi) = 0$. The peak or maximum elevation, $z = 1.05$ km, is located in the center of the domain where $x = y = \pi/2$ because $\sin(\pi/2) = 1$ is a maximum.

The black lines are contours of constant height. In Figure 6-2 they are drawn at 0.1 km = 100 m intervals – 0 m, 100 m, etc. The spacing of contour lines tells the slope of the land. Where contour lines are close, height changes in a short horizontal distance, d*s*, so the slope, d*z*/d*s* is large.

The shortest route to the summit is straight up the mountain but it may be too steep. That is why trails often snake back and forth, and why they are called snake paths. This increases trail length but makes the hike easier. Mountain roads are also often built as snake paths. The red trail in Figure 6-2 is a snake path.

By this point you may have asked, “Where is the math in this?” Here it is!

Problem: Find the average slope starting from $(x_i, y_i) = (0, 0)$ and ending at $x_f = \pi/2$!
Warning: This is a trick question because I didn’t give y_f. Because I didn’t, you can say, “If $y_f = 0$ then $z_f = 0$ and $\Delta z/\Delta x = 0$, but if $y_f = \pi/2$ that places us at the peak so $z_f = 1.05$ km and $\Delta z = 1.05$ km and $\Delta z/\Delta x = 1.05/1.507 = 0.67$.

The mathematical point is that once there are 2 or more independent variables, the slope or the derivative with respect to any one independent variable depends on the path and only has a unique value when the path is specified or constrained.

6.3 Total and Partial Derivatives and Differentials

The total derivative is the rate of change of a function when the path (i. e., the relation between all the independent variables) is specified.

The partial derivative of a function with respect to one independent variable is the function's rate of change when all the other independent variables are held constant.

For example, if a function, f(x, y), such as height of the land, depends on x and y, then the partial derivative of f(x, y) with respect to x when y is held constant is written as,

$$\left(\frac{\partial f}{\partial x}\right)_y$$

The symbol for partial derivatives, ∂, is like a **d** that has lost its tail, and so, is only partial. The subscripts indicate the independent variable(s) being held constant when the derivative is taken. Sometimes we get lazy and forget both the parentheses and the subscript, and assume that everyone else knows a partial derivative keeps **all** other independent variables constant. Partial derivatives are calculated the same way ordinary derivatives are calculated, with the one exception that, for the moment, all the other independent variables are treated as constants. For example, if f(x, y) = $3x^2y$ then the partial derivatives are given by

$$f(x,y) = 3x^2y \quad \Rightarrow \quad \left(\frac{\partial f}{\partial x}\right)_y = 6xy \quad \& \quad \left(\frac{\partial f}{\partial y}\right)_x = 3x^2$$

Next, let's talk about **differentials**. The total differential is the total change of a function when the start and end points of all independent variables are specified. It is the sum of all products of the change of each independent variable times the partial derivatives of the function with respect to that independent variable. The path between the endpoints won't change the total differential. (Crudely, it doesn't care how we get to be millionaires – the end justifies the means!) For the function, f(x, y), the total differential is,

$$\boxed{\mathrm{d}f(x,y) = \left(\frac{\partial f}{\partial x}\right)_y \mathrm{d}x + \left(\frac{\partial f}{\partial y}\right)_x \mathrm{d}y} \qquad \text{Eqn. 6-2}$$

The total differential looks something like a combination of the product rule and the chain rule of derivatives. Let's test Eqn. 6-2 with two simple problems.

Problem: Show that Eqn. 6-2 is correct for the function,

$$z(x,y) = x^2 + 2y \qquad \text{Eqn. 6-3}$$

Solution: Calculate the change of z from (x, y) to ($x + \Delta x$, $y + \Delta y$). If the changes in x and y are finite rather than infinitesimal, we use Δz, Δx, and Δy in place of dz, dx and dy.

$$z(x+\Delta x, y+\Delta y) - z(x,y) = \Delta z = (x+\Delta x)^2 + 2(y+\Delta y) - \left[x^2 + 2y\right]$$

Expanding, canceling, and recalling that $(\Delta x)^2 \ll \Delta x$ *when* $\Delta x \ll 1$ leads to

$$\Delta z = 2x\Delta x + 2\Delta y + (\Delta x)^2 \approx 2x\Delta x + 2\Delta y$$

This means that Eqn. 6-2 works because it matches the partial derivatives of Eqn. 6-3,

$$\left(\frac{\partial z}{\partial x}\right)_y = 2x \qquad \left(\frac{\partial z}{\partial y}\right)_x = 2$$

Problem: Show that Eqn. 6-2 is correct for the trickier function,

$$w(x, y) = x^2 y \qquad \text{Eqn. 6-4}$$

Solution: Eqn. 6-4 trickier because *x* and *y* are tied together in a single term and not merely added separately. Again, calculate the change of *w* from (*x*, *y*) to (*x* + Δ*x*, *y* + Δ*y*).

$$w(x+\Delta x, y+\Delta y) - w(x, y) = \Delta w = (x+\Delta x)^2 (y+\Delta y) - x^2 y$$

Once again, expanding and canceling where possible leads to

$$\Delta w = 2xy\Delta x + x^2 \Delta y + 2x\Delta x\Delta y + (\Delta x)^2 \Delta y$$

There are 4 terms on the right; two are first order in Δ*x* or Δ*y*, $(\Delta x)^2$ is 2nd order, and $(\Delta x)^2\Delta y$ is 3rd order. Again, I am sure that you remember from Chapters 1 and 2 that higher orders of tiny terms can be neglected as Δ*x* → 0 and Δ*y* → 0. Then Δ*w* simplifies to

$$\Delta w \approx 2xy\Delta x + x^2 \Delta y$$

Again, Eqn. 6-2 works because this matches the partial derivatives of *w* from Eqn. 6-4,

$$\left(\frac{\partial w}{\partial x}\right)_y = 2xy \qquad \left(\frac{\partial w}{\partial y}\right)_x = x^2$$

Total (Exact) Derivatives

When a total differential is divided by d*x* or d*s* or d*t*, it becomes a total derivative.

$$\boxed{\frac{\mathrm{d}f(x, y)}{\mathrm{d}s} = \left(\frac{\partial f}{\partial x}\right)_y \frac{\mathrm{d}x}{\mathrm{d}s} + \left(\frac{\partial f}{\partial y}\right)_x \frac{\mathrm{d}y}{\mathrm{d}s}} \qquad \text{Eqn. 6-5}$$

To get a better picture of what this means, we will calculate the slope A: directly up Double Sine Mountain and, B: on the snake path. But because this is a tough problem we will do a simpler problem first.

Problem: Calculate the slope (total derivative) d*z*/d*s* from Eqn. 6-3, $z = x^2 + 2y$, along a line *x* = 3*y* - 3 at point (*x*, *y*) = (3, 2) ….. (Isn't that a mouthful!)
Solution: Even this simpler problem takes several steps.
Step #1: Substitute Eqn. 6-3 into Eqn. 6-5 (with *z* = f) and take the partial derivatives of *z*.

$$\frac{\mathrm{d}z}{\mathrm{d}s} = 2x\frac{\mathrm{d}x}{\mathrm{d}s} + 2\frac{\mathrm{d}y}{\mathrm{d}s}$$

Step #2: Specify the path or direction to take its derivative. In other words, specify the relation between d*x* and d*y*. Finding the slope along the line *y* = 3*x* - 3 yields,

$$y = 3x - 3 \quad \Rightarrow \quad \mathrm{d}y = 3\mathrm{d}x$$

Step #3: Use Pythagoras, d*s* = $[(\mathrm{d}x)^2 + (\mathrm{d}y)^2]^{1/2}$ to solve for d*x*/d*s* and d*y*/d*s*.

$$(\mathrm{d}s)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2 = (\mathrm{d}x)^2 + (3\mathrm{d}x)^2 = 10(\mathrm{d}x)^2 \quad \Rightarrow \quad \frac{\mathrm{d}x}{\mathrm{d}s} = \frac{1}{\sqrt{10}} \quad \frac{\mathrm{d}y}{\mathrm{d}s} = \frac{3}{\sqrt{10}}$$

Step #4: (the final step) Substitute these values into the total derivative and evaluate at the given values of $x = 3$ and $y = 2$.

$$\frac{\mathrm{d}z}{\mathrm{d}s} = \frac{2x}{\sqrt{10}} + \frac{12}{\sqrt{10}} = \frac{18}{\sqrt{10}} \approx 4.11$$

That process wasn't too bad, was it? So now, let's do the tougher problem, which is only tougher because the equations are longer.

Problem: Calculate the slope **A**: directly up and, **B**: on the snake path of Double Sine Mountain at point $(x, y) = (\pi/4, \pi/2)$.

Solution A: If we go straight up the mountain starting from and ending at $x = \pi/2$, then $\sin(x) = \sin(\pi/2) = 1$. In that case, $\mathrm{d}x = 0$ so $\mathrm{d}s = \mathrm{d}y$, (so $\mathrm{d}y/\mathrm{d}s = 1$). Thus, the slope is,

$$\frac{\mathrm{d}z}{\mathrm{d}s} = \left(\frac{\partial z}{\partial x}\right)_y \frac{\mathrm{d}x}{\mathrm{d}s} + \left(\frac{\partial z}{\partial y}\right)_x \frac{\mathrm{d}y}{\mathrm{d}s} = \left(\frac{\partial z}{\partial y}\right)_x = 1.05\sin\left(\frac{\pi}{2}\right)\cos(y) = 1.05\cos\left(\frac{\pi}{4}\right) \approx 0.742$$

Solution B: We must use the equation of the snake path in order to determine its slope. The equation that I used to construct the snake path in Figure 6-2 is,

$$x = \frac{\pi}{2} + 0.4\left(1 - \frac{2y}{\pi}\right)\sin(32y) \qquad \text{Eqn. 6-6}$$

There is good news and bad news. The good news is that the procedure for calculating the slope along the snake path uses the same steps as before. Also, one long term drops out because it is multiplied by 0. The bad news is that the equations are obnoxiously long.

Step #1: Write the total derivative for Eqn. 6-1, the height of Double Sine Mountain.

$$\frac{\mathrm{d}z}{\mathrm{d}s} = \left(\frac{\partial z}{\partial x}\right)_y \frac{\mathrm{d}x}{\mathrm{d}s} + \left(\frac{\partial z}{\partial y}\right)_x \frac{\mathrm{d}y}{\mathrm{d}s} = 1.05\left[\cos(x)\sin(y)\frac{\mathrm{d}x}{\mathrm{d}s} + \sin(x)\cos(y)\frac{\mathrm{d}y}{\mathrm{d}s}\right]$$

Step #2: Substitute $y = \pi/4$ and $x = \pi/2$. (Note that $\cos(\pi/2) = 0$.)

$$\frac{\mathrm{d}z}{\mathrm{d}s} = 1.05\left[\cos\left(\frac{\pi}{2}\right)\sin\left(\frac{\pi}{4}\right)\frac{\mathrm{d}x}{\mathrm{d}s} + \sin\left(\frac{\pi}{2}\right)\cos\left(\frac{\pi}{4}\right)\frac{\mathrm{d}y}{\mathrm{d}s}\right] = \frac{1.05}{\sqrt{2}}\frac{\mathrm{d}y}{\mathrm{d}s}$$

Step #3: Differentiate Eqn. 6-6 of the snake path to relate $\mathrm{d}x$ and $\mathrm{d}y$.

$$\mathrm{d}x = 0.4\left[32\left(1 - \frac{2y}{\pi}\right)\cos(32y) - \frac{2}{\pi}\sin(32y)\right]\mathrm{d}y$$

Step #4: Calculate $\mathrm{d}y/\mathrm{d}s$ using $(\mathrm{d}s)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2$ to eliminate $\mathrm{d}x$. (Luckily, we don't have to calculate $\mathrm{d}x/\mathrm{d}s$ because at $x = \pi/2$ it is multiplied by $\cos(\pi/2) = 0$ in Step #2.) Thus,

$$\frac{\mathrm{d}y}{\mathrm{d}s} = \frac{1}{\sqrt{1 + \left\{0.4\left[32\left(1 - \frac{2y}{\pi}\right)\cos(32y) - \frac{2}{\pi}\sin(32y)\right]\right\}^2}}$$

Step #5: (the final step!) Substitute the equation for dy/ds from Step #4 into the equation for dz/ds from Step #2 and evaluate for y = π/4 and x = π/2 along the snake path. The slope then reduces to

$$\frac{\mathrm{d}z}{\mathrm{d}s} = \cos\left(\frac{\pi}{4}\right)\frac{1.05}{\sqrt{1+\{0.4[16]\}^2}} = \frac{1}{\sqrt{2}}\frac{1.05}{\sqrt{1+6.4^2}} \approx 0.115$$

This is much smaller than the slope straight up Double Sine Mountain at the same point = 1.05sin(π/4) ≈ 0.742. The snake path is much longer but its slope is much gentler.

6.4 Travelling Waves and Total Derivatives

So far, we have either written the sine wave as sin(kx), which describes a wave in space but not in time, or as sin(kt) as we did for spring motion, Eqn. 5-24 for a wave that changes in time but not space. But most waves vary and move over space AND time. So, now it is time to give some space to moving waves.

It is easy to understand moving waves when we see them. Unfortunately, animation in a book is not yet possible (except by flipping pages or throwing the book). To view wave motion, watch moving waves on the web or better yet, see or create them. You can create moving waves by taking a string or a slinky and shaking it back and forth or by partly filling your sink or bathtub with water and disturbing the surface with a finger.

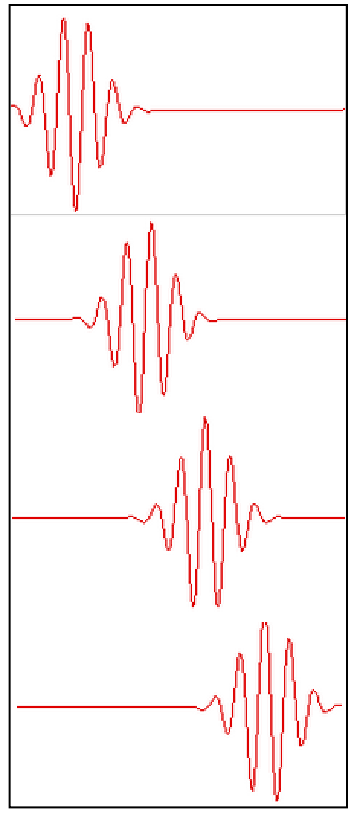


Figure 6-3 A wave packet moves from left to right from the top to the bottom frame.

Figure 6-3 does just about as much as I can do here to illustrate moving waves. It shows a group of waves or wave packet that is located further right in each of 4 successive frames from top to bottom. To visualize wave motion, either scan down or scroll up. If you look carefully, you might also notice that the waves in the packet change shape. This occurs because individual waves move faster than the packet, growing as they approach the center and then shrinking toward the right side of the packet. This is a characteristic of waves in deep water because such waves travel faster when they are longer (such waves are called dispersive), but it is not a characteristic of sound waves, which all travel with a speed that does not change with wavelength. Light waves are slightly dispersive – longer waves (red) travel slightly faster than shorter (violet) – through various substances such as water, glass, or diamond, which explains why they can and do split into a spectrum of different colors (or wavelengths).

As for the mathematics, we will only consider nondispersive waves that consist of simple sines and cosines. The almost universal letter used for wave speed is ***c*** which stands for **c**elerity. The letter, ***v***, which stands for **v**elocity, is not used for wave speed because waves often move rapidly through media that remain relatively still and for which $v \approx 0$. Thus, light waves move through a prism that does not move and even water waves move at a different speed than the water. If the ocean moved with ocean waves, boats could cross the ocean without sails or engines but the ocean would inundate the land.

How can we represent a moving wave mathematically? Velocity or celerity equals the change of distance divided by the time interval. Then, if A: wave speed, *c*, is constant and, B: time starts from $t = 0$, *c* is,

$$c = \frac{\Delta x}{\Delta t} = \frac{x - x_0}{t} \qquad \text{Eqn. 6-7}$$

Solving Eqn. 6-7 for the constant, x_0, shows that $x - ct$ is also constant.

$$x_0 = x - ct = const$$

But the sine of a constant is also constant, so,

$$\boxed{z = \sin(k(x - ct)) = \sin(kx_0) = const} \qquad \text{Eqn. 6-8}$$

This is not true for any random pair (*x, t*), but it is true when values of (*x, t*) are required to obey Eqn. 6-7. If $x - ct$ is constant then so is *z*, and the total differential, d*z* is zero. So if *z* is constant, the total differential, partial differential game yields,

$$\mathrm{d}z = \mathrm{d}[\sin(k(x - ct))] = \frac{\partial z}{\partial x}\mathrm{d}x + \frac{\partial z}{\partial t}\mathrm{d}t = 0$$

There are at least two paths to take now. One is to solve for d*x*/d*t*. This leads to a really curious and completely counterintuitive general result about the relation between total and partial derivatives, namely, whenever the independent variables are restricted to change so that the dependent variable remains constant (i. e, they have lost some freedom), then

$$\boxed{\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\left(\frac{\partial z}{\partial t}\right)_x}{\left(\frac{\partial z}{\partial x}\right)_t}} \qquad \text{Eqn. 6-9}$$

The first time I saw Eqn. 6-9 I was sure the minus sign was wrong. Indeed, every single time I look at that minus sign in Eqn. 6-9, it strikes me as strange. If you could freely cancel both of the ∂z's on the right hand side of Eqn. 6-9 you would be left with d*x*/d*t* = -$\partial x/\partial t$ and similar incorrect cancelations (of ∂x with d*x* and ∂t with d*t*) would ultimately reduce it to 1 = -1. But the two ∂z's do not cancel because each partial derivative comes with a different restriction, and gives the change of *z* in a different direction, so despite appearances, the two ∂z's are not the same, and partial and total differentials are not the same.

We can use Eqn. 6-8 to show that the minus sign in Eqn. 6-9 is true. The partial derivatives are,

$$\frac{\partial z}{\partial x} = k\cos(k(x - ct)) \qquad \frac{\partial z}{\partial t} = -kc\cos(k(x - ct)) \qquad \text{Eqn. 6-10}$$

Substituting both equations of Eqn. 6-10 into Eqn. 6-9 yields

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{-kc\cos(k(x - ct))}{k\cos(k(x - ct))} = c$$

The two minus signs cancel. What remains is a mathematical statement of the fact that z does not change if we move with the wave at its speed, dx/dt = c.

6.5 The First Order Wave Equation

The two partial derivatives of Eqn. 6-10 are related by the first order wave equation,

$$\boxed{\frac{\partial z}{\partial t} = -c\frac{\partial z}{\partial x}} \qquad \text{Eqn. 6-11}$$

We have done something astounding and really unusual here! The usual thing is to define a problem and derive a governing equation first and then undergo the long struggle to find solutions. Here, we went backwards – we began with a solution, namely a moving wave, sin[$k(x - ct)$] and proceeded to find its governing equation.

The first order wave equation, Eqn. 6-11 is used to describe water waves and traffic flow, but is limited in that it can only describe waves moving in one direction. We get a more general wave equation if we take the second partial derivatives of z in Eqn. 6-8,

$$\frac{\partial^2 z}{\partial x^2} = -k^2 \sin(k(x-ct)) \qquad \frac{\partial^2 z}{\partial t^2} = -(kc)^2 \sin(k(x-ct))$$

Relating these produces the...

6.6 The Classical Wave Equation

$$\boxed{\frac{\partial^2 z}{\partial t^2} = c^2 \frac{\partial^2 z}{\partial x^2}} \qquad \text{Eqn. 6-12}$$

Eqn. 6-12 is the 1-Dimensional Classical Wave Equation. (The 3-D version is Eqn. 6-61.) Because c^2 includes both $\pm c$ it allows waves to travel in the + or - x direction. It governs a host of wave phenomena in electromagnetism, fluids, acoustics, and elasticity. Each phenomenon has its own equation for wave speed based on the governing physics.

Electromagnetic waves include visible light, ultraviolet radiation, microwaves, and radio waves. All travel at the speed of light, which (in the vacuum of empty space) is given by

$$\boxed{c = \sqrt{\frac{1}{\varepsilon_0 \mu_0}} \approx 2.998(10)^8 \ m \cdot s^{-1}} \qquad \text{Eqn. 6-13}$$

The strange symbols, $\varepsilon_0 = 8.85(10)^{-12}$ and $\mu_0 = 1.26(10)^{-6}$ are the electric constant and the magnetic constant respectively. (Confession: I looked up Eqn. 6-13 because I forgot it.)

Water waves are called shallow when their wavelengths are much longer than the depth, h, of the water. This includes tsunamis. All shallow water waves travel with speed,

$$\boxed{c_{shallow} = \sqrt{gh}} \qquad \text{Eqn. 6-14}$$

Here, $g \approx 10$ m·s^{-2} is the magnitude of the acceleration of gravity.

Problem: Calculate c for the Tsunami of 26 December 2004 and how long it took to reach

Sri Lanka from its source off the west coast of Sumatra, Indonesia.
Data: The distance is just about 1600 km and the mean depth of the ocean is $h \approx 4800$ m.
Solution:

$$c = \sqrt{gh} \approx \sqrt{10 \cdot 4800} \approx 220\ \mathrm{m \cdot s^{-1}}$$

$$c = \frac{\Delta s}{\Delta t} \quad \Rightarrow \quad \Delta t = \frac{\Delta s}{c} \approx \frac{1.6(10)^6\ \mathrm{m}}{220\ \mathrm{m \cdot s^{-1}}} \approx 7270\ \mathrm{s}$$

A wave speed of 220 $\mathrm{m \cdot s^{-1}}$ is almost 500 mph. Tsunamis move almost as fast as jets. The 2004 Tsunami needed only 2 hours to cross the Indian Ocean. This would have provided more than enough time to sound sirens and evacuate the shores of Sri Lanka if there had been a tsunami warning network across the Indian Ocean. But there wasn't.

Unfortunately, no warning system can ever help much in places near a tsunami's source, as in Banda Aceh, Indonesia because there is almost no time – just minutes or seconds – to evacuate, and, if the land is level, the water will proceed rapidly inland, as we saw in videos of the 07 April 2011 Tsunami of Tōhoku, Japan.

For deep water waves the depth is greater than the wavelength, λ.[1] Their speed is,

$$\boxed{c_{deep} = \sqrt{\frac{g}{k}} = \sqrt{\frac{g\lambda}{2\pi}}} \qquad \text{Eqn. 6-15}$$

Deep water waves are dispersive – the longer the wavelength, λ the faster the speed (as in Fig. 6-3). Storms generate a jumble of waves of all lengths. The longer waves move out faster and strike distant shores first. As the storm approaches and the time to evacuate gets shorter, so does the wavelength. Centuries before there were weather forecasts, the long waves, or **swell**, gave beach dwellers and surfers many hours of warning of approaching hurricanes and typhoons. Movies that show people sailing out in calm waters in the hours before a hurricane arrives are baloney. They are designed solely to scare us.

Problem: Calculate the speed of a waves with wavelength, λ = 500 m and λ = 100 m in the deep ocean and calculate how long it takes such waves to move 1000 km.
Solution: For 500 m long waves:

$$c = \sqrt{\frac{g\lambda}{2\pi}} = \sqrt{\frac{10(500)}{2\pi}} \approx 28.2\ m \cdot s^{-1} \qquad \Delta t = \frac{\Delta s}{c} \approx \frac{1000\ km}{28.2\ m \cdot s^{-1}} \approx 35{,}450\ s \approx 9.85\ h$$

I'll let you calculate the speed and elapsed time for λ = 100 m long waves! The answer: wave speed is $c \approx 12.6$ $\mathrm{m \cdot s^{-1}}$ and the elapsed time for the waves to arrive is $\Delta t \approx 22.0$ h.

Warning: Waves give less warning time the faster a storm approaches. In the extreme

[1] The equations for shallow and deep water waves are approximations of the general formula for small amplitude water waves, $c^2 = [g\lambda/2\pi]\cdot\tanh[2\pi h/\lambda]$ for small and large values of $[2\pi h/\lambda]$ respectively. You knew I would find another important application of the hyperbolic tangent!

case that the storm approaches as fast as the longest waves travel, the waves will give no warning time at all. Usually though, the longest waves move much faster than storms do.

6.7 Diffusion and the Classical Diffusion Equation

Many of Nature's physical phenomena can be classified in two grand categories, namely waves and diffusion. Waves exemplify the springy, eternally youthful aspect of Nature while diffusion exemplifies the degrading, aging aspect of Nature. Waves propagate or transmit contrasts; diffusion spreads and wears down contrasts. In principle, Waves are Capitalist, highlighting differences; Diffusion is Communist, seeking uniformity.

Heat conduction through material that does not move is a form of diffusion. Imagine walking barefoot across the sand on a sunny summer afternoon. The sand at the surface burns your feet but dig your feet several inches down and the sand there is much cooler – little of the day's heat has been conducted down to it. If you return to the beach several hours after sunset, you will notice that the sand at the surface has become cool, but now that you know about conduction dig your feet several inches down and you will not be surprised that the sand is warmer because some of the day's heat has finally reached it.

Our old friend, Jean Baptiste Joseph Fourier improved upon a finding of Isaac Newton and demonstrated that heat is transmitted from hot to cold regions in solids at a rate or *Flux Density*, F_q with units of $W \cdot m^{-2}$ (Watts per square meter) that is proportional to the temperature gradient, i. e, to the partial derivative of T with respect to distance at any fixed time, t. In one dimension this takes the form,

$$\boxed{F_q = -k\frac{\partial T}{\partial x}} \qquad \text{Eqn. 6-16}$$

Alert #1: Temperature is always represented by Upper Case ***T***, time by lower case ***t***!
Alert #2: Don't confuse Flux Density with the flux capacitor in ***Back to the Future***.

The identical law of diffusion was discovered by Adolf Fick for smelly gases diffusing through air, by Georg Ohm for electrical currents through a wire with resistance (this led to the dominance of AC over DC), and by Henry Darcy for flow of water through sand or porous rock (this is important for wells and water loss from reservoirs).

The minus sign in Eqn. 6-16 indicates that heat travels in the direction opposite the gradient, from regions of high to low T. Thermal conductivity, k, is a physical property of a material (See Table 6-1). It ranges from about 2200 for diamond to less than 0.01 for silica aerogel. Note that silver transmits heat more than 12000 times faster than Styrofoam. Silver transmits heat so rapidly that holding a silver cup of boiling coffee will burn your hands severely, while a Styrofoam cup of boiling coffee is easy to hold.

Material	k ($J \cdot m^{-1} °C^{-1} s^{-1}$)
Diamond	2200
Silver	430
Glass	1.2
Wood	0.1
Styrofoam	0.033
Silica Aerogel	0.01

Table 6-1 Some values of Thermal conductivity.

Problem: Calculate the Flux Density of heat through a silver cup 2.5 mm thick and onto your skin if the cup is holding boiling water.
Information: The temperature of boiling water (at sea level) is 100°C and the temperature of your body is 37°C (and your skin is cooler than body temperature).
Solution: Simply (yes, simply) substitute all values into Eqn. 6-16.

$$F_q = -k\frac{\partial T}{\partial x} \approx -430\frac{(100-37)}{0.0025} \approx -10^7 \text{ W}\cdot\text{m}^{-2}$$

When you look at this result, think only of its magnitude. (The minus sign tells the direction of heat flow, which everyone knows is from the boiling water to your skin.) That is about 10,000 times faster than the heat flux of direct sunlight, roughly 1000 W·m^{-2}, which only makes your hand feel warm. It is even far greater than sunlight focused by a magnifying glass on a small spot of your hand, so it is certainly enough to burn your hand. Holding a Styrofoam cup of boiling water feels OK because Styrofoam conducts heat at only 1/12,000th as fast as silver, or about 800 W·m^{-2}, which is similar to direct sunlight.

Heat Conduction and the Classical Diffusion Equation

Eqn. 6-16 tells the flow rate of heat, but not if or how fast T is changing. T increases when heat flows into a fixed volume faster than it flows out (neglecting heat sources in the volume). This is clearly the case in the cold center of box A of Figure 6-4 because heat flows towards it from each warm side.

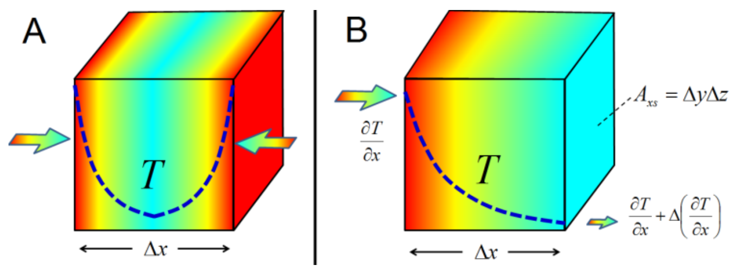


Figure 6-4 T and heat flow in tiny boxes. Both boxes are warming. A: Heat flows from both sides towards cold core. B: Heat flows toward right throughout the box but faster on left side, so it accumulates in box.

Box B shows a subtler situation. Heat flows from the warm left side toward the cold right side, but because the temperature gradient is larger on the left side than on the right, heat enters the left side faster than it leaves the right. This region is therefore warming (assuming k is constant).

The rate of temperature change at any point is equal to the partial derivative, $(\partial T/\partial t)_x$. (Don't forget capital T stands for Temperature, small t, for time.) Figure 6-4B shows how it is calculated. Heat flows into the box's left side at a rate,

$$A_{xs}F_q = -kA_{xs}\frac{\partial T}{\partial x}$$

The gradient on the right side of the volume, a distance Δx away, differs from the gradient on the left side by an amount, $[\partial/\partial x(\partial T/\partial x)]\Delta x$, so heat flows out at the rate given by

$$A_{xs}(F_q + \Delta F_q) = -kA_{xs}\left[\frac{\partial T}{\partial x} + \frac{\partial}{\partial x}\left(\frac{\partial T}{\partial x}\right)\Delta x\right]$$

The tiny box acts like any system. Its heating rate, $\partial Q/\partial t$ equals the inflow rate of heat minus the outflow rate of heat or,

$$\left(\frac{\partial Q}{\partial t}\right)_x = \Delta F_q = -kA_{xs}\frac{\partial T}{\partial x} + kA_{xs}\left[\frac{\partial T}{\partial x} + \frac{\partial}{\partial x}\left(\frac{\partial T}{\partial x}\right)\Delta x\right] = kA_{xs}\frac{\partial^2 T}{\partial x^2}\Delta x \qquad \text{Eqn. 6-17}$$

Now we must relate the heat change to the temperature change. (A cup of warm water will warm a mouse but would not do much for an elephant.) To do this we must invoke a relatively simple equation from Physics with a very fancy name - the ***First Law of Thermodynamics***. When all heat is used to raise T - heat can also melt solids, boil water, or run engines - the ***First Law*** states that the change of heat of an object equals the change of T times the mass, m, times the specific heat capacity, c of the material. Then, taking the partial derivative with respect to time to get a heating rate, the First Law is,

$$\left(\frac{\partial Q}{\partial t}\right)_x = cm\left(\frac{\partial T}{\partial t}\right)_x \qquad \text{Eqn. 6-18}$$

Specific heat capacity, c, is the amount of heat needed to raise T of 1 kg of a material by 1°C, and is a property of the material (see Table 6-2). For example, it takes 18 times more heat to increase T of 1 kg of water by 1°C than 1 kg of silver by 1°C. Thus, c is 18 times greater for water than for silver.

Material	c ($J{\cdot}kg^{-1}{}^{\circ}C^{-1}$)
Water	4186
Air	1004
Silver	233
Glass	840

Table 6-2 Specific heat capacities, c, of several materials

The form of Eqn. 6-18 we need occurs when we set mass $m = \rho V$, where ρ is density and V is volume, and then set $V = A_{xs}\Delta x$, where $A_{xs} = \Delta y{\cdot}\Delta z$ is the cross section area of the volume and Δx is the thickness. This transforms Eqn. 6-18 to

$$\left(\frac{\partial Q}{\partial t}\right)_x = c\rho A_{xs}\Delta x\left(\frac{\partial T}{\partial t}\right)_x \qquad \text{Eqn. 6-19}$$

Combining Eqn. 6-17 with Eqn. 6-19 leads to

$$c\rho A_{xs}\Delta x\left(\frac{\partial T}{\partial t}\right)_x = kA_{xs}\left(\frac{\partial^2 T}{\partial x^2}\right)_t \Delta x \qquad \text{Eqn. 6-20}$$

Rearranging and defining diffusivity, $D \equiv k/(c\rho)$ and cancelling A_{xs} and Δx yields...

The Classical Diffusion Equation

$$\boxed{\left(\frac{\partial T}{\partial t}\right)_x = \frac{k}{c\rho}\left(\frac{\partial^2 T}{\partial x^2}\right)_t = D\left(\frac{\partial^2 T}{\partial x^2}\right)_t} \qquad \text{Eqn. 6-21}$$

This is the 1-Dimensional form of the Classical Diffusion Equation. (The 3-D version is Eqn. 6-62.) The dotted curves in Figure 6-5 show how diffusion changes the profile of T with time. At time $t = 0$ there is a narrow region of high T near the center of a material (red dotted curve). With time the maximum T decreases as the warm region spreads. This occurs rapidly at first and ever more slowly as t continues until eventually the entire universe is slightly warmer.

The classical diffusion equation and its solutions are beautiful (at least to physicists and mathematicians) but it is a Partial Differential Equation (PDE), so how in the world are we going to solve it? With the classical wave equation (also a PDE) we were extremely lucky to start with the solution. For the classical diffusion equation we have no such luck.

I won't go through the complete solution, because it takes us way beyond the first year of Calculus, but I now show the remarkable and simple trick called **Separation of Variables**, which makes it much easier to solve many important PDE's. That trick has the same spirit as the trick for solving Ordinary Differential Equations (ODE's) with constant coefficients, namely magically transforming them to much simpler equations.

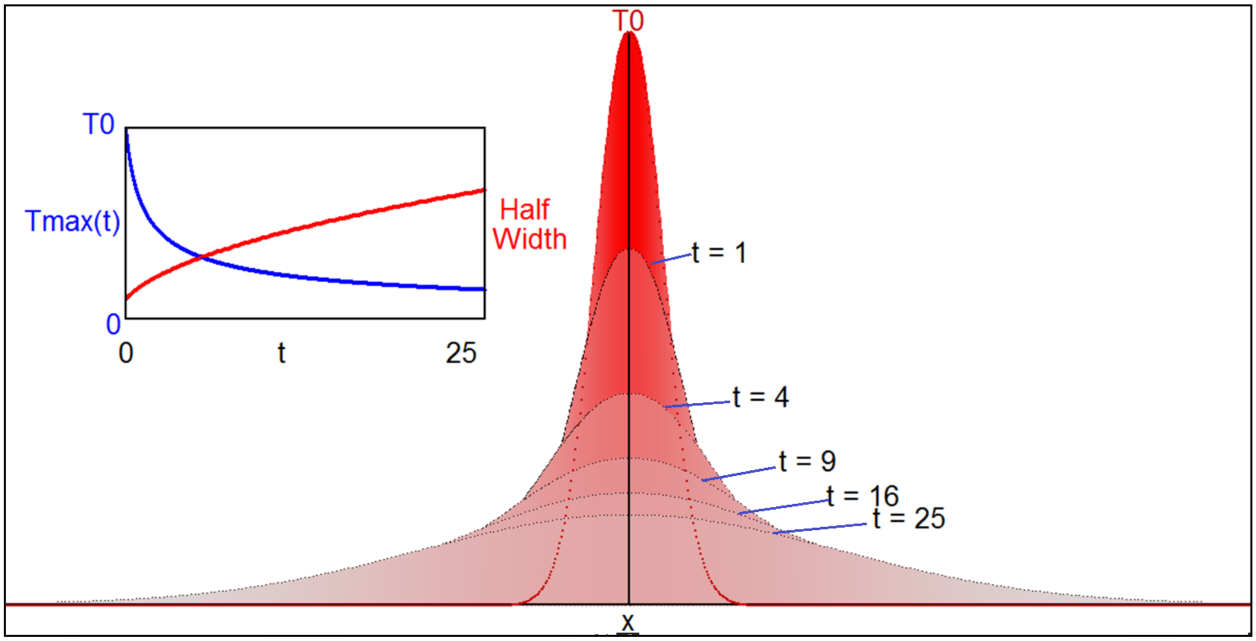


Figure 6-5 Evolution of the temperature profile with time due to diffusion. The inset graph shows how T_{max} decreases with time as the width of the warm region (red curve) increases.

6.8 Separation of Variables: A Trick for Solving Partial Differential Equations

The trick for solving the classical diffusion equation is to assume that *T* (which is a function of both *t* and *x*) is a product of a function of *t* only times a function of *x* only, or

$$T(x,t) = G(t) \cdot H(x) \qquad \text{Eqn. 6-22}$$

Substituting Eqn. 6-22 into the classical diffusion equation, Eqn. 6-21 leads to,

$$H(x)\frac{\mathrm{d}G(t)}{\mathrm{d}t} = D \cdot G(t)\frac{\mathrm{d}^2 H(x)}{\mathrm{d}x^2} \qquad \text{Eqn. 6-23}$$

The next step gives the second great critical insight. Group all functions of *t* on the left side and all functions of *x* on the right side. This separates the variables and converts the PDE (which is impossible for humans to solve) into two ODE's, which we can solve.

$$\frac{1}{G(t)}\frac{\mathrm{d}G(t)}{\mathrm{d}t} = C = \frac{D}{H(x)}\frac{\mathrm{d}^2 H(x)}{\mathrm{d}x^2} \qquad \text{Eqn. 6-24}$$

The key insight to recognize about Eqn. 6-24 is, since both sides always remain equal but vary independently, neither side can vary at all and so, must equal some constant, *C*.

Thus, Eqn. 6-24 is essentially two ODE's with constant coefficients. This, in turn, means that both ODE's have exponential solutions and can be transformed to algebra equations, which are even easier to solve. How sweet it is!

Solving each of these ODE's separately gives the general solution,

$$G(t) = G(0)e^{Ct} \qquad H(x) = Ae^{\sqrt{\frac{C}{D}}x} + Be^{-\sqrt{\frac{C}{D}}x} \qquad \text{Eqn. 6-25}$$

Mathematically, we call it quits at this point, but here are a few hints of how to continue. First, since diffusion causes *T* to level out and spread with time from any initial peak region, the time behavior, G(t) should be a <u>negative</u> exponential. But when *C* is negative, the square root, $(C/D)^{1/2}$ in the exponential of H(*x*) is an imaginary number. This means it consists of sines and cosines. And, how do we fit any arbitrary function using sines and cosines? With Fourier Series! So, solutions to many heat conduction problems involve Fourier Series, which is why Fourier was the first to solve them. Enough said!

6.9 Distances along Curves: Arc Length

It is easy to calculate distance along a straight line. When the line is horizontal or vertical it is easiest of all. If the line is diagonal it is a little tougher because then we need Pythagoras. But for straight lines we don't even need math – all we need is a ruler.

Finding distance along curves is tougher. Just try using a ruler around a curve. You probably would get a close estimate after a few tries if the ruler doesn't slip and if the curve is not too complex. But this clearly is a job for Calculus.

I only do two problems here. The first is easy. Find the length of the arc of a circle! Since we already used the definition of the radian (Section 0.5) and then found the value of π (Section 1.3) we only have a teeny, weeny bit more to do, namely rewrite Eqn. 0-20. Arc length of a circle equals the radius times the central angle of the arc.

$$\boxed{L_{\text{circle}} = r\Delta\theta = r(\theta_1 - \theta_2)} \qquad \text{Eqn. 6-26}$$

That was so easy it is barely a problem at all. Now for the tough one!

Problem: Calculate the length of the parabola $y = x^2$ from $x = 0$ to $x = 1$.
Solution: Arc length is the integral of elements of length, $\mathrm{d}s = (\mathrm{d}x^2 + \mathrm{d}y^2)^{1/2}$ based on Pythagoras. Reworking slightly, a general form for arc length is the integral,

$$\boxed{L = \int \mathrm{d}s = \int \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2} = \int_{x=0}^{1} \left[\sqrt{1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\right] \mathrm{d}x} \qquad \text{Eqn. 6-27}$$

For the parabola $y = x^2$, the derivative, $\mathrm{d}y/\mathrm{d}x = 2x$, so Eqn. 6-27 becomes,

$$L_{y=x^2} = \int_{x=0}^{1} \sqrt{1 + 4x^2}\,\mathrm{d}x = 2\int_{x=0}^{1} \left[\left(\frac{1}{2}\right)^2 + x^2\right]^{0.5} \mathrm{d}x \qquad \text{Eqn. 6-28}$$

Realizing Eqn. 6-28 is so difficult and important, I solved it in Chapter 3 as Eqn. 3-31.

Here, a = 0.5. All terms equal 0 when x = 0, so what remains are the terms when x = 1,

$$L=\left[\sqrt{1^2+0.5^2}\right]\cdot 1+0.5^2\ln\left[\frac{\sqrt{1^2+.5^2}+1}{.5}\right]\approx 1.479$$

Figure 6-6 shows that $L \approx 1.48$ makes sense because the parabola is longer than the straight diagonal line ($L = \sqrt{2} \approx 1.41$) but shorter than the arc of 1/4th of a circle ($L = \pi/2 \approx 1.57$) and certainly, two sides of the square (L = 2).

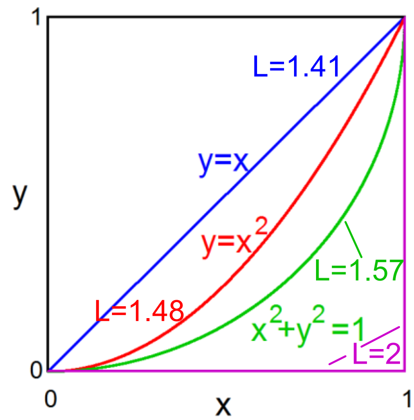


Figure 6-6 Arc lengths of parabola (red), diagonal line (blue), circular arc (green) and edge of square (purple).

6.10 Curvature of Curves

Perhaps rapidly spinning rides in the amusement parks excite you, but they nauseate me, so I am glad to watch others have a great time getting sick. One ride starts off like a Merry-Go-Round, but once it spins fast enough, it turns on its axis vertically and becomes more like a Ferris Wheel, with people glued to the inside of the wheel. The centripetal acceleration is so large that everyone is kept safely in place provided the ride keeps spinning rapidly.

This is one example of rotation, which is ubiquitous in nature and in engines. Almost all engines rotate or have spinning parts. The Earth undergoes several spinning motions simultaneously. It rotates on its axis roughly once a day, and revolves around the Sun once a year as the Sun itself revolves around the Milky Way Galaxy. I could go on but you get the idea.

Rotation and spinning are linked to curvature. Make a sharp turn in a rapidly moving automobile and everyone gets thrust to the outside of the curve. Curvature is a simple notion and has a simple definition. Curvature, K is the change of direction, θ (shown in Figures 0-19 and 0-20) divided by the distance traversed, or,

$$\boxed{K \equiv \frac{\mathrm{d}\theta}{\mathrm{d}s}} \qquad \text{Eqn. 6-29}$$

A circle is the only 2-D curve with constant curvature. The equation for curvature of a circle is ridiculously simple. For a circle of radius, r, arc length, $s = r\theta$, Eqn. 6-29 becomes,

$$\boxed{K_{\text{circle}}=\frac{\mathrm{d}\theta}{\mathrm{d}(r\theta)}=\frac{1}{r}\frac{\mathrm{d}\theta}{\mathrm{d}\theta}=\frac{1}{r}} \qquad \text{Eqn. 6-30}$$

Because $K = 1/r$, you can fit a circle to any point of any curve. Figure 6-7 shows a road where traffic moves in the direction of the arrows. The road curves clockwise first and then much more sharply counterclockwise. The red circles match the curvature of the road. Curvature is greatest where the matching circle is smallest.) Curvature, K = 0 at the inflection point, a point where $d^2y/dx^2 = 0$ (noted by *). There, the road has just stopped curving clockwise and is about to curve counterclockwise.

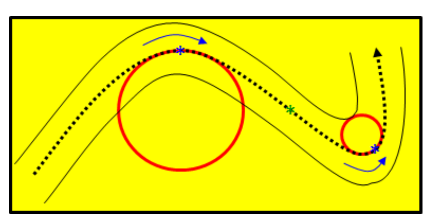


Figure 6-7 Curvature of a road matched by red circles. It is 0 at the inflection point (green *).

Whenever curvature varies on a curve, Eqn. 6-29 may look pretty, but it is useless for most curves until it gets a makeover, which involves expressing angle, θ (from the *x* axis) in terms of the slope, d*y*/d*x*. Figure 0-20 (which is valid for a circle of any size such as radius, d*r* with height d*y* and lateral distance d*x*) plus your knowledge of trigonometry imply

$$\tan(\theta) = \frac{\mathrm{d}y}{\mathrm{d}x} \quad \Rightarrow \quad \theta = \tan^{-1}\left[\frac{\mathrm{d}y}{\mathrm{d}x}\right]$$

Substituting this into Eqn. 6-29, then the Chain Rule and Pythgoragas yields,

$$K = \frac{\mathrm{d}\theta}{\mathrm{d}s} = \frac{\mathrm{d}\theta}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}s} = \frac{\mathrm{d}\left(\tan^{-1}\left[\frac{\mathrm{d}y}{\mathrm{d}x}\right]\right)}{\mathrm{d}x}\left[1+\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right]^{-\frac{1}{2}}$$

Recalling the derivative of $\tan^{-1}(u)$ (Eqn. 2-37) and setting u = d*y*/d*x* leads to

$$\frac{\mathrm{d}\tan^{-1}[u]}{\mathrm{d}x} = \frac{1}{1+u^2}\frac{\mathrm{d}u}{\mathrm{d}x} \quad \Rightarrow \quad K = \left[1+\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right]^{-1}\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{\mathrm{d}y}{\mathrm{d}x}\right]\left[1+\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right]^{-\frac{1}{2}}$$

The Equation of Curvature

$$\boxed{K = \frac{\left[\frac{\mathrm{d}^2 y}{\mathrm{d}x^2}\right]}{\left[1+\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\right]^{1.5}}} \qquad \text{Eqn. 6-31}$$

Eqn. 6-31 shows that curvature is proportional but not equal to the second derivative, i. e., to the rate of change of the slope. At an inflection point (*K* = 0) the curve is straight and curvature changes from left to right or from concave to convex and the 2nd derivative is zero.

Now, we can calculate curvature of any curve.

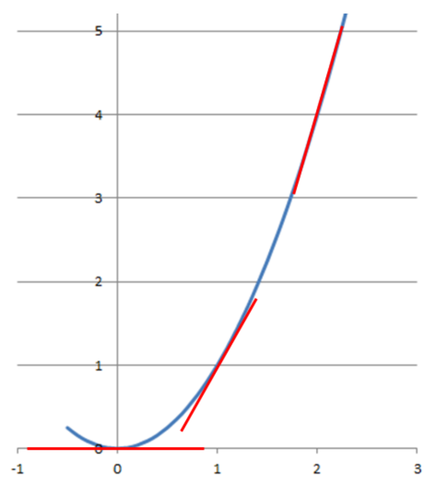


Figure 6-8 Mismatches of a parabola (blue) and straight lines (red) show curvature.

Problem: Calculate the curvature of the parabola, $y = x^2$ when *x* = 0, 1, and 2.

Solution: For the parabola, Eqn. 6-31 for curvature simplifies to

$$K_{y=x^2} = \frac{2}{\left[1+(2x)^2\right]^{1.5}}$$

Thus, curvature decreases as x^2 increases. $K = 2$ at $x = 0$, $K = 2/5^{1.5} \approx 0.179$ at $x = 1$, and, $K = 2/17^{1.5} \approx 0.0285$ at $x = 2$. The mismatches between the straight red line segments and the blue parabola in Figure 6-8 show that the parabola's curvature is visibly smaller at $x = 1$ than at $x = 0$ and is so small at $x = 2$, that it is barely detectable.

Curves, the Planets and Centripetal Acceleration

The planets, Earth included, ~~circle~~ orbit the Sun in great ellipses. Did you ever stop to think why they do not fall into the Sun? After all, they are drawn towards the Sun by gravitational attraction and there is no repelling force.

Each planet's motion prevents it from falling into the Sun. Gravitation draws it toward the Sun. By Newton's Law, it accelerates toward the Sun. The planets' acceleration toward the center of a curve is called **Centripetal Acceleration**, which is given by,

$$\boxed{a_{\text{cen}} = \frac{v^2}{r} = Kv^2} \qquad \text{Eqn. 6-32}$$

We know enough Calculus and Physics to derive this important equation. Figure 6-9 illustrates the situation. An object moves around a circle at constant speed, $v = r\cdot d\theta/dt$. Starting at $y = 0$ it moves directly up the circle toward the top of the page. After time, Δt it has moved a distance $r\Delta\theta$, where its component of motion toward the left is,

$$\Delta u = -v\sin(\Delta\theta)$$

When $\Delta\theta$ is small, $\sin(\Delta\theta) \approx \Delta\theta$ (Eqn. 2-2). So,

$$\Delta u = -v\cdot\sin(\Delta\theta) \approx -v\cdot\Delta\theta$$

To verify Eqn. 6-32, divide both sides by Δt and replace $\Delta\theta/\Delta t$ with v/r.

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -v\lim_{\Delta t\to 0}\frac{\Delta\theta}{\Delta t} = -\frac{v^2}{r}$$

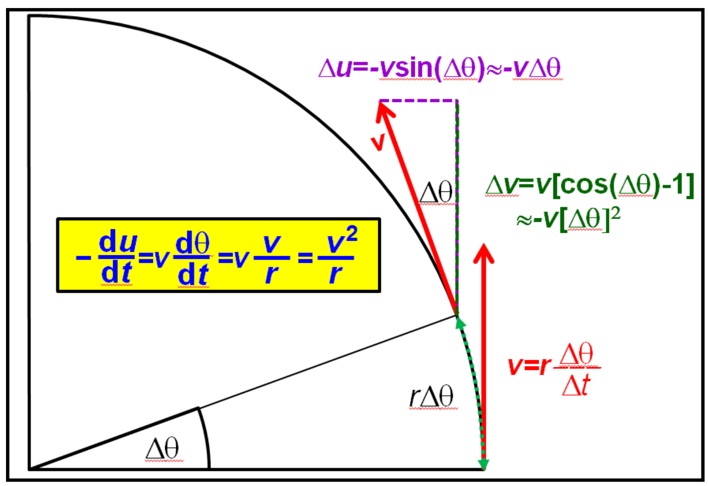


Figure 6-9 Circular motion, centripetal acceleration.

If you are troubled about the minus sign, "fugget about it"! The minus sign tells the acceleration's direction, which is towards the left or the negative x direction in this case.

By the way, Eqn. 2-3 shows that acceleration in the y direction at $y = 0$ is 0 because $dv/dt \approx v[\cos(\Delta\theta) - 1]/\Delta t \approx -v(\Delta\theta)^2/\Delta t = -v[\Delta\theta/\Delta t]\cdot\Delta\theta \to 0$ as $\Delta\theta \to 0$. Thus, Centripetal acceleration always points toward the center of the circle or curve.

Problem: A vomit ride moves at $v = 10\ \mathrm{m\cdot s^{-1}}$ around a radius of 5 m. Is this enough to keep riders who are not belted in from falling off?

Solution: Riders will not fall off so long as the centripetal acceleration at the top of the ride (which points downward) is greater than gravity, $g \approx 10\ \mathrm{m\cdot s^{-2}}$. In this case,

$$a_{\text{cen}} = \frac{v^2}{r} = \frac{10^2}{5} = 20\ \mathrm{m\cdot s^{-1}} > g$$

Since $a_{\text{cen}} > g$ (and is directed downward at the top of the ride) the riders will not fall off!

The opposite problem is to ask why Earth's rotation does not fling us out into space, as the ride would if we were not constrained. Early critics of the idea that the Earth revolves around the Sun (and rotates every day) thought that if the Earth rotated we might fall off because of centrifugal force. Let us test that hypothesis. We will go to the equator where we are furthest from the Earth's axis of rotation and where centripetal acceleration is largest. Then $r = R_E = 6.37(10)^6$ m, $v = ds/dt = 2\pi r/t_d$ and $t_d = 86400$ s.

Problem: Why doesn't Earth's rotation fling us out into space?
Solution: We must show that its centripetal acceleration is much smaller than the acceleration of gravity. Substitute the information in the paragraph above about Earth's radius and the velocity due to its rotation into Eqn. 6-32 for centripetal acceleration.

$$a_{\text{cen}} = \frac{v^2}{r} = \frac{1}{R_E}\left(\frac{2\pi R_E}{t_d}\right)^2 \frac{1}{R_E} = \frac{4\pi^2 R_E}{t_d{}^2} = \frac{4\pi^2(6.37)(10)^6}{8.64^2(10)^8} = 0.0337\ \text{m}\cdot\text{s}^{-2}$$

This means that a_{cen} at the Equator is only about 1/300th of g ($\approx 10\ \text{m}\cdot\text{s}^{-2}$). Therefore someone who weighs 300 pounds will be one pound lighter at the Equator than at the North or South Pole where there is no centripetal acceleration due to Earth's rotation. Clearly Earth's rotation provides neither the best way to lose weight nor enough centripetal acceleration to fling objects on its surface out into space.

6.11 Spherical Coordinates: Volumes, Areas and Distances

Planets and stars are all almost perfectly spherical. Rotation bulges them slightly outward at their equators, and planets with solid surfaces, like Earth, also have mountains and valleys that make the surface rough, but when you see the Earth from space you cannot detect these tiny 'imperfections'.

Spherical coordinates provide the best way to describe events that occur on a sphere, such as waves produced by major earthquakes or impacts of asteroids. Spherical Coordinates, illustrated in Figure 6-10, may seem difficult and strange, but they are the natural coordinates for all of us on Earth. They consist of

1: radius, ***r***, distance from Earth's center.

2: longitude, θ, (sometimes λ) or angle east of a line running from the North Pole to South Pole passing through Greenwich, England.

3: latitude or colatitude, ϕ.

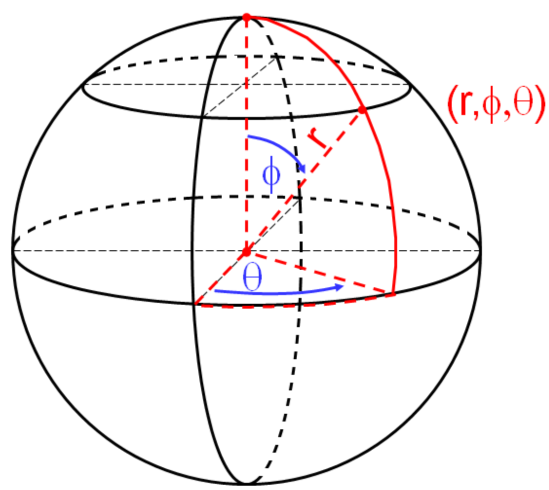


Figure 6-10 Spherical Coordinates (r, θ, ϕ).

Latitude is the angle north of the Equator. Colatitude is the angle from the North Pole. We are familiar with latitude, which extends from $-\pi/2$ (-90°) at the South Pole, to 0 at the Equator, to $+\pi/2$ (90°) at the North Pole. Colatitude starts with 0 at the North Pole and goes to π (180°) at the South Pole and is a bit simpler mathematically than latitude.

Figure 6-11 relates Spherical and Cartesian Coordinates. Distance from the origin in Cartesian Coordinates is based on doing Pythagoras twice and is $(x^2 + y^2 + z^2)^{½}$ but is simply r in Spherical Coordinates. Height above the origin (or equatorial plane) is z in

Cartesian Coordinates and $r{\cdot}\cos(\phi)$ in Spherical Coordinates. Solving for ϕ, yields $\phi = \cos^{-1}(z/r)$. In this book I use capital $R = r{\cdot}\sin(\phi)$ for horizontal distance from the origin (i. e, distance from Earth's Axis of Rotation, the line that passes through the Poles). The x and y components of R are, $x = R{\cdot}\cos(\theta) = r{\cdot}\sin(\phi){\cdot}\cos(\theta)$ and $y = R{\cdot}\sin(\theta) = r{\cdot}\sin(\phi){\cdot}\sin(\theta)$. Longitude, θ, is found by taking the inverse tangent of $y/x = \tan(\theta)$. Putting this all together, the equations for Cartesian Coordinates, x, y, and z in terms of Spherical Coordinates are,

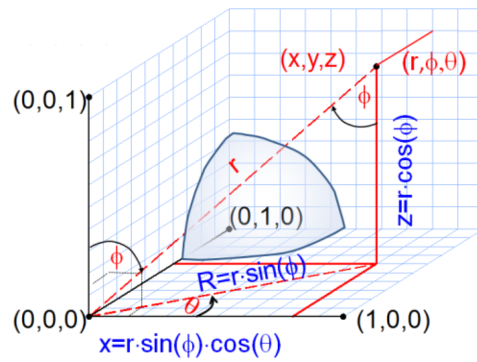


Figure 6-11 Relating Cartesian and Spherical Coordinates.

$$\boxed{\begin{aligned} x &= r\sin(\phi)\cos(\theta) \\ y &= r\sin(\phi)\sin(\theta) \\ z &= r\cos(\phi) \end{aligned}}$$ Eqn. 6-33

The equations for Spherical Coordinates r, ϕ, and θ in terms of Cartesian Coordinates are,

$$\boxed{\begin{aligned} r &= \sqrt{x^2+y^2+z^2} \\ \phi &= \cos^{-1}\left(\frac{z}{r}\right) \\ \theta &= \tan^{-1}\left(\frac{y}{x}\right) \end{aligned}}$$ Eqn. 6-34

Volume

Volume is length × width × height. In Cartesian Coordinates, this leads directly to a 3-D or triple integral for volume,

$$\boxed{V = \int\int\int \mathrm{dxdydz}}$$ Eqn. 6-35

Problem: Calculate the Volume of a Square Base Pyramid with base = b and height = h.
Solution: The width and length of the pyramid decrease linearly from b at z = 0 to 0 at $z = h$. Therefore, the limits for both x and y are, $x = y = b(1 - z/h)$. The triple integral becomes,

$$V = \int_0^h \int_0^{b(1-z/h)} \int_0^{b(1-z/h)} \mathrm{dxdydz} = \int_0^h b^2\left(1\text{-}\frac{z}{h}\right)^2 \mathrm{dz} = b^2\left(h - 2\cdot\frac{h^2}{2h} + \frac{h^3}{3h^2}\right) = \frac{b^2 h}{3}$$

Therefore, the volume of a square base pyramid is,

$$\boxed{V_{pyr} = \frac{1}{3}b^2 h}$$ Eqn. 6-36

You can see from Figure 6-12 that the volume of a square base pyramid is much smaller than the volume of the box enclosing it. Can you guess where the factor, ⅓ in Eqn. 6-36 comes from? Since both the pyramid's length and width decrease linearly with z, area of a horizontal slice of the pyramid decreases with z^2. Integrating thin slices of the pyramid's area over height is then like integrating z^2dz which produces the factor, ⅓.

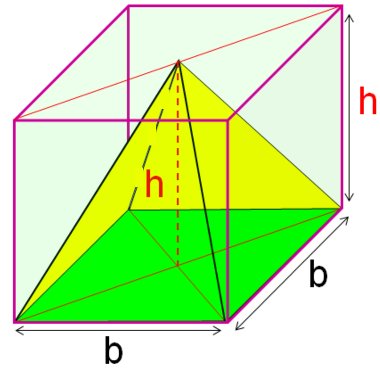


Figure 6-12 Square-base pyramid in a box.

There are at least two ways to derive the triple integral for volume in Spherical Coordinates. We could go through the long and painful process of using Eqn. 6-33 to transform the volume element, dV = dxdydz from Cartesian to Spherical Coordinates or we can use the less rigorous but much simpler, visually motivated technique of using the length elements in Spherical Coordinates. I hope you can see from Figure 6-13 and from your knowledge of arc length that the elements of length, ds_i in the r (radial), ϕ (latitudinal) and, θ (longitudinal) directions are,

$$\boxed{\begin{aligned} \mathrm{d}s_1 &= \mathrm{d}r \\ \mathrm{d}s_2 &= r\mathrm{d}\phi \\ \mathrm{d}s_3 &= r\sin(\phi)\mathrm{d}\theta \end{aligned}}$$

Eqn. 6-37

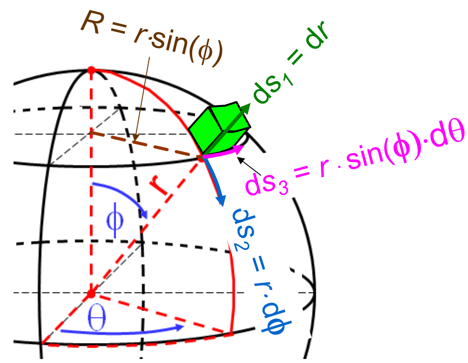


Figure 6-13 Elements of Length in Spherical Coordinates.

Eqn. 6-37 implies that the elements of surface area and volume in Spherical Coordinates are, dA = $r^2\sin(\phi)\mathrm{d}\theta\mathrm{d}\phi$ = ds_2·ds_3 and dV = $r^2\sin(\phi)\mathrm{d}\theta\mathrm{d}\phi\mathrm{d}r$ = ds_1·ds_2·ds_3.The integrals for surface area and volume in spherical coordinates are then,

$$\boxed{A = \int\int r^2\sin(\phi)\mathrm{d}\theta\,\mathrm{d}\phi}$$

Eqn. 6-38a

$$\boxed{V = \int\int\int r^2\sin(\phi)\mathrm{d}\theta\mathrm{d}\phi\mathrm{d}r}$$

Eqn. 6-38b

Problem: Calculate the volume of a sphere with radius, r.
Solution: For the entire sphere, radius goes from 0 to r, colatitude, ϕ, from 0 to π, and longitude, θ, from 0 to 2π. With these limits, Eqn. 6-38b becomes,

$$V_{sph} = \int_0^r\int_0^\pi\int_0^{2\pi} r^2\sin(\phi)\mathrm{d}\theta\mathrm{d}\phi\mathrm{d}r = 2\pi\int_0^r\int_0^\pi r^2\sin(\phi)\mathrm{d}\phi\mathrm{d}r = 2\pi[\cos(0)-\cos(\pi)]\int_0^r r^2\mathrm{d}r = 4\pi\frac{r^3}{3}$$

Volume of a Sphere!!!

$$\boxed{V_{sph} = \frac{4}{3}\pi r^3}$$

Eqn. 0-19

Remember this Equation from Chapter 0? Am I crazy or was deriving Eqn. 0-19 (from Eqn. 6-38b) too easy? Deriving the volume of sphere so simply (admittedly after the long setup) is one of the great, astounding results of Calculus! I think I need to catch my breath.

Growing Raindrops: Volume and Proportions

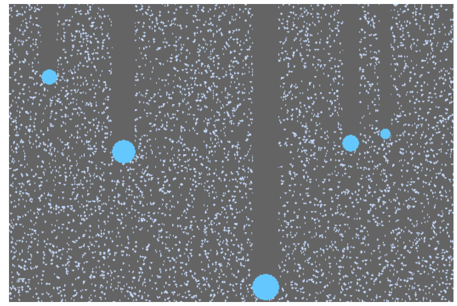


Figure 6-14 Trails of large drops collecting small droplets as they slide down a mirror.

Many clouds seem to float overhead without a single drop falling from them. Typical droplets in clouds are so tiny ($r_{cd} \approx 10$ μm = micrometers ≈ 0.0004"), that they fall so slowly (≈ 0.01 m·s^{-1}) they would need about 24 hours to fall 1 km. And, even if the tiny droplets do fall from the cloud, they evaporate so quickly they don't get far.

Drops must grow much larger to make rain. A typical raindrop is $r_{rain} \approx 1000$ μm (1 millimeter). It falls ≈ 7 m·s^{-1} ≈ 15 mph, so that it will fall 1 km in ≈ 2 minutes and easily reach the ground from a cloud.

How do cloud droplets grow? You can see the main process on a misty mirror or Figure 6-14. Any drop that grows large enough begins to slide down the mirror, leaving a clear wake where it collected all the tiny droplets in its path. So,

Problem: How many cloud droplets of $r_{cd} \approx 10$ μm must a raindrop collect to grow to $r_{rain} \approx 1000$ μm if the raindrop starts as a cloud droplet? Hint: This isn't even Calculus!

Solution: We must solve for the number of cloud droplets, N_{cd}, that have the same total volume as a single raindrop. This problem is easiest to solve using proportions. The radius of a typical raindrop is 100 times larger than the radius of a typical cloud droplet. But since volume (not radius) is the relevant quantity and volume is proportional to radius cubed (r^3),

$$N_{cd} \cdot V_{cd} = 1 \cdot V_{rain} = N_{cd}\frac{4}{3}\pi r_{cd}^3 = \frac{4}{3}\pi r_{rain}^3 \quad \Rightarrow N_{cd} = \left(\frac{r_{rain}}{r_{cd}}\right)^3 = (100)^3 = 10^6$$

A typical raindrop has about the same mass as 1 million cloud droplets. That's an awful lot of droplets to collect to reach raindrop size!

Mass of a Sphere

A somewhat more difficult integral occurs when we want to integrate any function, such as density, $\rho(r, \phi, \theta)$ that varies inside a sphere. To find the mass of the Earth, for example, we must integrate density over the volume. This is easiest when ρ depends only on r.

$$\boxed{M = \int\int\int \rho \cdot r^2 \sin(\phi)\,\mathrm{d}\theta\,\mathrm{d}\phi\,\mathrm{d}r} \qquad \text{Eqn. 6-39}$$

Problem: Calculate the mass of the Earth.

Information: Earth's radius is $r_E \approx 6.37(10)^6$ m. Density of rocks near Earth's surface, $\rho_E \approx$ 3000 kg·m^{-3} or about 3 times that of water. Mean density increases toward the center of the Earth, where $\rho_E \approx$ 13,000 kg·m^{-3} because materials are compressed by ever increasing pressure and because the central region, called the Core, consists largely of iron, which is much denser than all rocks. Density can therefore be approximated by two equations – one for the iron Core and one for the rocky Mantle that lies outside the Core.

$$\rho_E \approx \begin{matrix} 1.3(10)^4 - 10^{-3} r & \quad r < 3.5(10)^6 \\ 8(10)^3 - 7(10)^{-4} r & \quad r > 3.5(10)^6 \end{matrix}$$

Solution: Since density depends almost entirely on radius and is essentially independent of latitude or longitude, the integral for mass simplifies to,

$$M_E \approx 4\pi\left[\int_0^{3.5(10)^6}\left[1.3(10)^4 - 10^{-3}r\right]\cdot r^2\mathrm{d}r + \int_{3.5(10)^6}^{6.37(10)^6}\left[8(10)^3 - 7(10)^{-4}r\right]\cdot r^2\mathrm{d}r\right]$$

I'll let you solve this equation, which consists of 4 terms. The calculated answer is $5.88(10)^{24}$ kg, which is pretty close to the actual mass of the Earth, $M_E \approx 5.98(10)^{24}$ kg.

Surface Area of a Sphere

Surface Area of a sphere is the double integral at constant radius given by Eqn. 6-38a. We can use this to find the surface area of any part of the Earth or of any planet.

Problem: Derive Eqn. 0-18 for the Surface Area of a Sphere.
Solution: Eqn. 6-38a yields the amazingly simple and profound result that ***$A_{sph} = 4\pi r^2$***!

$$\boxed{A_{sph} = r^2\int_{\phi=0}^{\pi}\int_{\theta=0}^{2\pi}\sin(\phi)\mathrm{d}\theta\mathrm{d}\phi = 2\pi r^2\int_{\phi=0}^{\pi}\sin(\phi)\mathrm{d}\phi = 4\pi r^2} \qquad \text{Eqn. 0-18}$$

Problem: Calculate the surface area of the Earth and of the oceans.
Information: The oceans occupy 71% of the surface area of Earth.
Solution: This is *just* algebra! Apply Eqn. 0-18 using r_E to find the area of the Earth.

$$A_E = 4\pi r_E^2 \approx 4\pi\left[6.37(10)^6\right]^2 \approx 5.10(10)^{14}\ \mathrm{m}^2$$

Earth's surface area, $A_E \approx$ 510 million square km. Multiply by 0.71 to get A_{Oceans}.

$$A_{Oceans} \approx 0.71\cdot A_E \approx (0.71)(5.10)(10)^{14}\ \mathrm{m}^2 \approx 3.62(10)^{14}\ \mathrm{m}^2$$

Problem: If all the ice on Greenland and Antarctica melted, how much will sea level rise?
Information: The global volume of ice would melt to an extra ocean volume, $\Delta V_{\text{ice-water}} \approx 2.9(10)^{16}\ \mathrm{m}^3$ or about 29 million cubic km.
Solution: This too is *just* algebra. Since $\Delta V = A\Delta h$, solve for Δh.

$$\Delta h_{SL} = \frac{\Delta V_{ice-water}}{A_{Oceans}} \approx \frac{2.9(10)^{16}\ \mathrm{m}^3}{3.62(10)^{14}\ \mathrm{m}^2} \approx 80\ \mathrm{m}$$

If all the ice on Earth melted, sea level would rise just about 80 m or 254 feet.

Distance on a Sphere: The Great Circle is the Shortest Distance between Points

If your house is less than about 80 m above sea level you will want to know how far you must travel to get to dry ground once all the ice melts. You might look at a map and draw a straight line from your home to the destination. But on a sphere, a straight line is not the shortest distance between two points, especially for long distances. The shortest path between two points on a sphere is a Great Circle. You can find a Great Circle by stretching a string across a globe. Then you will see why flights between North America and Asia or Europe often travel quite near the North Pole, as in Figure 6-15 (even though I doubt there are direct flights between Moscow, Russia and Moscow Idaho).

If Z (for Zenith) is the angle between two points on a sphere, then Eqn. 6-26 gives the minimum distance on the Great Circle Route as $s_{\min} = rZ$. The cosine of the angle, Z is given by the long equation from spherical trigonometry

$$\boxed{\cos(Z) = \sin(\phi_1)\sin(\phi_2) + \cos(\phi_1)\cos(\phi_2)\cos(\theta_1 - \theta_2)} \qquad \text{Eqn. 6-40}$$

Thus, distance between two points on a sphere depends only on the latitudes, ϕ_1 and ϕ_2 and longitudes, θ_1 and θ_2 of the two points (as well as the radius of the Earth).

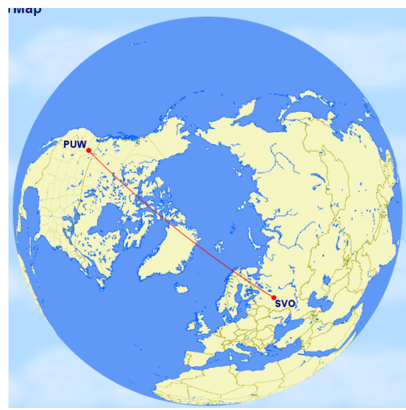


Figure 6-15 Great Circle Route from Idaho to Russia.

Warning: In Eqn. 6-40 ϕ is latitude, NOT colatitude.

Problem: Calculate the Distance between Moscow, Idaho and Moscow, Russia.
Information: The latitudes and longitudes for each Moscow are ($\phi_1 \approx 46.44° \approx 0.81$, $\theta_1 \approx -117.0° \approx -2.04$) and ($\phi_2 \approx 55.75° \approx 0.973$, $\theta_2 \approx 37.62° \approx 0.657$) respectively. The units are degrees and radians respectively.
Solution: Use Eqn. 6-45 to find angle, Z.

$$\cos(Z) \approx \sin(0.81)\sin(0.973) + \cos(0.81)\cos(0.973)\cos(-2.04 - 0.657) \approx 0.248$$

To find angle, Z, take the inverse cosine. This yields,

$$Z \approx 1.32 \text{ radians} \approx 75.6^\circ$$

Eqn. 6-26 shows that distance around a circle equals the radius times the angle of the arc, thus rZ, Thus, the distance between the two Moscow's is,

$$s_{\min} = r_E Z \approx 6370 \times 1.32 \approx 8407 \text{ km}$$

The two Moscow's are little more than 8400 km or 5200 miles apart, and the great circle route in Figure 6-15 crosses over northern Greenland, close to the North Pole.

6.12 Cylindrical Coordinates

When a pebble drops into water, circular waves emanate from the point of impact (Figure 6-16). The simplest way to describe the properties of the waves is to use cylindrical coordinates (Figure 6-17). The coordinates are (R, θ, z), where z is the height, R is the distance from the z axis and θ is the azimuthal angle or the angle from the x-axis.

Warning: I use capital R for Radius in cylindrical coordinates to distinguish it from r in spherical coordinates.

Figure 6-16 Circular waves emanating from a spash.

In cylindrical coordinates R is the horizontal distance from the z axis (sometimes the axis of rotation). In spherical coordinates, r is the 3-D distance from the origin. When ϕ is colatitude, we have already seen (from Figure 6-13) that the relation is,

$$R = r\sin(\phi)$$

The equations for x, y, z, in terms of Cylindrical Coordinates are,

$$\boxed{\begin{aligned} x &= R\cos(\theta) \\ y &= R\sin(\theta) \\ z &= z \end{aligned}} \qquad \text{Eqn. 6-41}$$

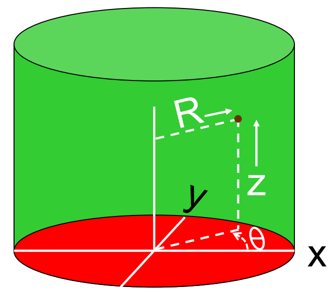


Figure 6-17 Cylindrical vs Cartesian Coordinates.

The equations for R, θ, z, in terms of Cartesian Coordinates are,

$$\boxed{\begin{aligned} R &= \sqrt{x^2 + y^2} \\ \theta &= \tan^{-1}\left(\frac{y}{x}\right) \\ z &= z \end{aligned}} \qquad \text{Eqn. 6-42}$$

Figure 6-17 also shows that the elements of length in Cylindrical Coordinates are

$$\boxed{\begin{aligned} \mathrm{d}s_1 &= \mathrm{d}R \\ \mathrm{d}s_2 &= R\mathrm{d}\theta \\ \mathrm{d}s_3 &= \mathrm{d}z \end{aligned}} \qquad \text{Eqn. 6-43}$$

Eqn. 6-48 leads directly to the integral for volume in cylindrical coordinates,

$$\boxed{V = \int\int\int \mathrm{d}s_1\mathrm{d}s_2\mathrm{d}s_3 = \int\int\int R\mathrm{d}\theta\mathrm{d}r\mathrm{d}z} \qquad \text{Eqn. 6-44}$$

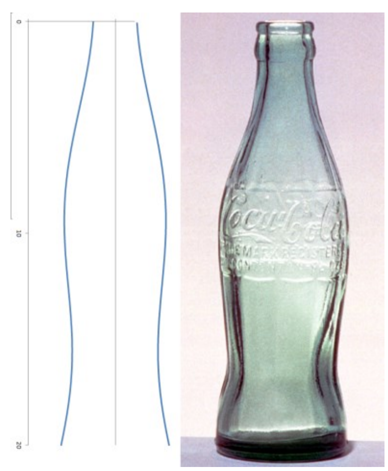


Figure 6-18 A function that fits inside of a Coke Bottle.

Problem: Calculate the volume of the classic 6.5 ounce (192 cm^3) Coca Cola bottle (Figure 6-18).

Information: The bottle is almost 20 cm high but the glass is thick so the inside is smaller. Furthermore, the coke was not filled to the top, and the bottom of the bottle was quite thick so I guessed that the integral of z extends from about 2 to 18.5 cm and the inside radius of the bottle of Figure 6-18) is,

$$R \approx 1.4 + 0.06z - 0.4\cos(0.12\pi z)$$

Solution:

$$V_{sph} = \int_2^{18.5}\int_0^{r}\int_0^{2\pi} R\mathrm{d}\theta\mathrm{d}r\mathrm{d}z = \pi\int_2^{18.5} R^2\mathrm{d}z = \pi\int_2^{18.5}\left[1.4 + 0.06z - 0.4\cos(0.12\pi z)\right]^2\mathrm{d}z$$

This integral is simple but long enough to drive you to drink (and not Coke). What we have to do is square the expression in brackets and then integrate each of the six resulting terms. I leave it for you to do. I did it numerically and got 215 cm^3, which is 10% high. I think that I underestimated the thickness of the glass. Thicker glass leaves less space inside. The high cost of manufacturing these bottles caused their demise; now they cost far more as collector's items, as is the case with so much other worthless old trash. Indeed, *The Gods Must Be Crazy*!

6.13 Vectors: Size Matters - and so does Direction!

When you want to navigate around your world, whether it's two or three dimensions, size matters, but so does direction. After all there is a great difference between having \$100 and owing \$100, or between walking 100 miles to the east and 100 miles to the west, or between gaining 100 pounds and losing 100 pounds! Because both size (technically called magnitude) and direction are important in so many situations and for so many phenomena, mathematicians, said, "Let there be Vectors!"

Vectors are quantities that contain both magnitude and direction but no fixed starting point. As a result, they are often depicted as arrows. A long arrow has a large magnitude. And, of course, there is a big, big difference between an arrow that points up and one that points down. Vectors point, but never specify where they start or end.

Quantities that only involve magnitude are called **scalars**. Up until now, we have been working exclusively with scalars. All along you may have noticed that *scalars are written in regular italics*. By contrast, **vectors are written upright in boldface**.

Vectors constitute a host of quantities in nature including displacement (distance moved in a specified direction), velocity, acceleration, force and electric, magnetic, and gravitational fields. Several more complicated quantities, such as gradients or slopes, fluxes, and vorticity (spin of motion) are also vectors.

The direction of a plane or surface is defined by its normal vector, **n,** which stands at right angles to the plane or surface. The arrows in Figure 6-19 show **n** for a point on a sphere and its tangent plane.

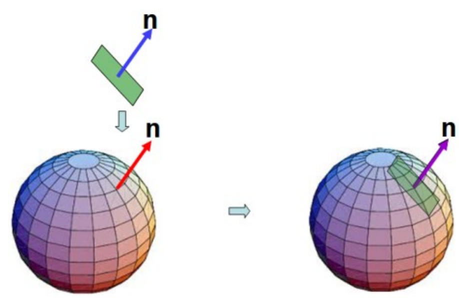


Figure 6-19 The green plane has normal vector, **n**, tangent to the sphere at the red arrow's base.

When a vector, **V** is placed on a 3-dimensional Cartesian graph, as in Figure 6-20, you can see that 1: its components have magnitudes V_x, V_y and V_z, and, 2: The ||magnitude|| of the vector, given by the Pythagorean Theorem, is written in standard form inside double vertical lines, so that,

$$\boxed{\|\mathbf{V}\| = \sqrt{V_x^{\,2} + V_y^{\,2} + V_z^{\,2}}} \qquad \text{Eqn. 6-45}$$

It is also standard to write vectors in terms of the component magnitudes, V_x, V_y, and V_z, and the **unit vectors**, **i**, **j**, and **k**, which have magnitude 1 and point in the positive *x*, *y*, and *z* directions respectively.

$$\boxed{\mathbf{V} = V_x\mathbf{i} + V_y\mathbf{j} + V_z\mathbf{k}} \qquad \text{Eqn. 6-46}$$

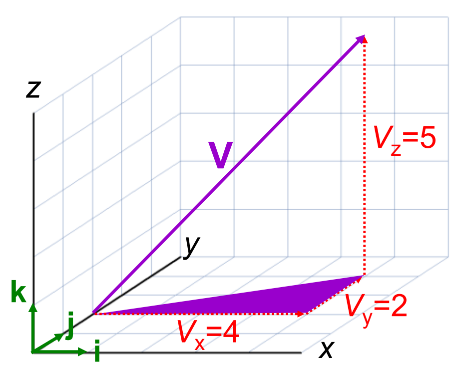


Figure 6-20 Components of **V** and unit vectors, **i**, **j**, **k** in Cartesian Coordinates, *x*, *y*, *z*.

The value of **V** (the purple arrow) in Figure 6-20 is,

$$\mathbf{V} = 4\mathbf{i} + 2\mathbf{j} + 5\mathbf{k}$$

Vector Math

Vector math is shorthand for describing processes and phenomena that involve direction. For example, vectors provide a compact way to write Newton's Equations of Motion and Maxwell's Equations of Electricity and Magnetism. Vectors are also great at proving general results such as the Law of Cosines, Eqn. 0-10. (Do you remember it? That ought to force you to look back.) But, **Warning:** While vectors are great at getting some general results, good old scalars are often better at solving particular problems.

Vector Math can be subdivided into Vector Algebra and Vector Calculus. Table 6-3 gives a list of some of the most common procedures that involve vectors. An example (I will soon illuminate) is that the scalar or dot product tells us how power from a solar panel depends on how it is aimed.

Vector Algebra	
1. Scalar Multiplication	$c\mathbf{V}=cV_x\mathbf{i} + cV_y\mathbf{j} + cV_z\mathbf{k}$
2. Vector Addition	$\mathbf{A} + \mathbf{B} = (A_x + B_x)\mathbf{i} + (A_y + B_y)\mathbf{j} + (A_z + B_z)\mathbf{k}$
3. Scalar or Dot Product	$\mathbf{A} \cdot \mathbf{B} = A_xB_x + A_yB_y + A_zB_z$
4. Vector or Cross Product	$\mathbf{A} \times \mathbf{B} = (A_yB_z - A_zB_y)\mathbf{i} + (A_zB_x - A_xB_z)\mathbf{j} +(A_xB_y - A_yB_x)\mathbf{k}$
5. Matrix Multiplication	**AB** = not in this book!
Vector Calculus	
6. Gradient	$\nabla f = \mathbf{i}(\partial f/\partial x) + \mathbf{j}(\partial f/\partial y) + \mathbf{k}(\partial f/\partial z)$
7. Divergence	$\nabla \cdot \mathbf{V} = \partial V_x/\partial x + \partial V_y/\partial y + \partial V_z/\partial z$
8. Curl or Vorticity	$\nabla \times \mathbf{V} = (\partial V_z/\partial y - \partial V_y/\partial z)\mathbf{i} + (\partial V_x/\partial z - \partial V_z/\partial x)\mathbf{j} + (\partial V_y/\partial x - \partial V_x/\partial y)\mathbf{k}$
9. Laplacian	$\nabla \cdot \nabla f = \partial^2 f/\partial x^2 + \partial^2 f/\partial y^2 + \partial^2 f/\partial z^2$

Table 6-3 Overview of Vector Math Operations. I treat all but #5 in this book.

1. Product of a Vector and a Scalar

To multiply a vector by a scalar (such as c), simply multiply each component of the vector by that scalar, as Eqn. 6-52 shows.

$$\boxed{c\mathbf{V} = cV_x\mathbf{i} + cV_y\mathbf{j} + cV_z\mathbf{k}} \qquad \text{Eqn. 6-47}$$

For example, two identical locomotives will pull a train down a track with twice the force of one so, each component of the original force will double.

2. Vector Addition

Adding vectors is just as easy. Since the starting point for vectors doesn't matter, the easiest way to add vectors is to move them head to tail like a line of elephants, as in Figure 6-21. The only difference is that vectors keep their own direction, while elephants tend to follow the leader and move along a more-or-less straight line.

Figure 6-21 Vectors, like elephants, are added by moving them so that they line up head to tail. The vector sum is **r** = **a** + **b** + **c** + **d.** The diagram at far right shows how to move vectors head to tail and then find their scalar components with Cartesian Coordinates.

The rule for adding vectors was designed for adding forces. Decompose each vector into a magnitude or change of *x*, *y*, and *z*. Then add each of the component magnitudes or changes separately, as in Eqn. 6-48 and Table 6-4.

$$\boxed{\mathbf{A}+\mathbf{B}=(A_x+B_x)\mathbf{i}+(A_y+B_y)\mathbf{j}+(A_z+B_z)\mathbf{k}} \qquad \text{Eqn. 6-48}$$

Name	Δx	Δy
a	6	-1
b	4	3
c	1	-3
d	-1	-3
Sum		
r	10	-4

Table 6-4 Adding vectors of Figure 6-21.

Problem: Find the Vector Sum of the vectors of Figure 6-21.
Solution: By placing the vectors on a Cartesian graph, we see that their values are those given in Table 6-4. (Note that these vectors are restricted to the *x*-*y* plane.) Then simply apply Eqn. 6-48 and add each component to get the resultant, **r**.

Let's use vectors to move around the perimeter of a triangle, as in Figure 6-22. We treat each of the three sides as a vector so the triangle's three vectors are, **a**, **b**, and **c**.

When we move around the perimeter of a triangle we return to the point we started. In effect, we have gone nowhere. In terms of vectors, this means, **a** + **b** + **c** = **0**! We will use this result to prove the law of cosines right after defining the scalar or dot product.

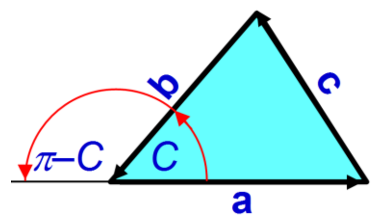


Figure 6-22 The vectors of a triangle add to zero.

3. The Scalar or Dot Product

Here are three scenarios to use the dot product. 1: You want to divert water from a river into an irrigation channel. You open a gate that projects into the river, like an open door projects into a room. The flow rate or flux of water the gate intercepts depends on the angle of the gate to the direction of flow. 2: You install solar panels on the roof. The power the panels deliver depends on the angle of the panels to the direction of the sunbeams. 3. You build an electric motor or generator. (They contain wire loops). The power depends on the angle between the loop and the magnetic field.

In these three cases the dot product enables us to calculate the flux of water, solar energy flux and magnetic flux respectively. Consider the solar panel. To maximize its power, it must face the Sun squarely. In vector terms the normal, **n** to the panel must be parallel to the

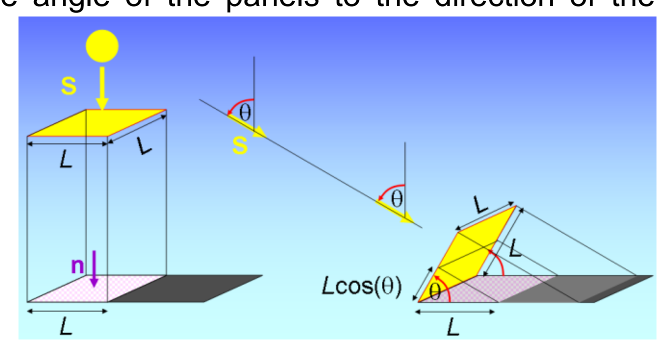


Figure 6-23 The cosine law for a solar panel.

Sun's rays, which reach Earth with magnitude and direction, **S**. Imagine a (pink) square solar panel, with sides of length, *L* on a level roof, as in Figure 6-23. The area of the panel can be written as **A** = L^2**n**, because it is a vector that points in the **n** direction, which for this panel is vertical. When the Sun is overhead, the area of sunlight the panel intercepts is L^2. Since direct solar irradiance, ||**S**|| ≈ 1360 W·m^{-2} (Watts per square meter) minus obstructions in the atmosphere, the power the solar panel will receive is, $P_{max} \approx 1360L^2$ W (Watts).

However, as the Sun goes down it strikes the horizontal panel obliquely, and a part of the sunbeam will miss the panel. The right diagram in Figure 6-23 shows that when the Sun is at angle, θ from the vertical, the area of the sunbeam striking the horizontal panel is $L^2\cos(\theta)$, so the power the panel will receive is $P \approx 1360L^2\cos(\theta)$ W.

This is exactly what the scalar or dot product does. In terms of the dot product, the power received by solar panel with area L^2 is,

$$P = \mathbf{S} \cdot \mathbf{A} \approx 1360L^2\cos(\theta)$$

This result is also known as the cosine law and it is quite general. Another important application of the cosine law and dot product involves Work, which is defined in Physics as Force times the component of distance traversed in the direction of the Force, or d*W* = **F** · d**s** = *F*d*s*cos(θ). The scalar or dot product is the component of the product of the magnitudes of two vectors in the direction of either vector. It is given by Eqn. 6-49,

$$\boxed{\mathbf{a} \cdot \mathbf{b} = ab\cos(\theta) = \left(a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}\right) \cdot \left(b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}\right) = a_xb_x + a_yb_y + a_zb_z} \qquad \text{Eqn. 6-49}$$

The magnitude of the dot product of two vectors is a maximum when two vectors are parallel and zero when they are perpendicular. The scalar or dot product therefore uses the fact that the dot products of parallel unit vectors equal 1 or, **i** · **i** = **j** · **j** = **k** · **k** = 1 and the dot products of perpendicular unit vectors equal 0, or, **i** · **j** = **i** · **k** = **j** · **k** = 0.

Problem: Use the Dot Product to prove the Law of Cosines
Solution: Recall that traversing the sides of a triangle means that the 3 vectors add to 0. Rearranging and taking the dot product of **c** with itself, yields,

$$\mathbf{a} + \mathbf{b} + \mathbf{c} = 0 \quad \Rightarrow \mathbf{c} = -(\mathbf{a} + \mathbf{b})$$
$$\mathbf{c} \cdot \mathbf{c} = (\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) \quad \Rightarrow \mathbf{a} \cdot \mathbf{a} + \mathbf{b} \cdot \mathbf{b} + 2\mathbf{a} \cdot \mathbf{b}$$

The dot product of a vector with itself is simply the magnitude squared. Then using Eqn. 6-49 and noting that the angle between vectors **a** and **b** in Figure 6-22 is not *C* but π - C [and cos(π - *C*) = -cos(*C*)], yields the Law of Cosines (Eqn. 0-10).

$$\mathbf{c} \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{a} + \mathbf{b} \cdot \mathbf{b} + 2\mathbf{a} \cdot \mathbf{b} \quad \Rightarrow \quad c^2 = a^2 + b^2 - 2ab\cos(C)$$

4. The Vector or Cross Product.

The Vector or Cross Product is designed to describe some of Nature's most curious, even mystifying behaviors. I will present three scenarios that use the vector or cross product, 1: Leverage and Torque, 2: Rotating Systems, 3: Electric and Magnetic Fields.

First, loosen a rusty bolt. This can be a wrenching experience. You must turn the wrench - it is useless to push it into or out of the bolt. A given force is much more effective with a longer wrench because it has greater leverage (see Figure 6-24). Thus, the relevant quantity is not force but the torque. Torque, τ is the vector equal to the cross product of the force and the lever arm, i. e., the radius vector that extends from the pivot point to the point the force is applied, as in Eqn. 6-50.

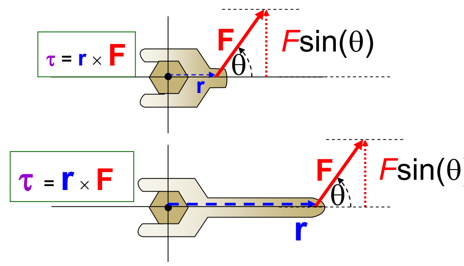


Figure 6-24. Torque, **r** × **F**, is larger the longer the wrench.

$$\boxed{\boldsymbol{\tau} = \mathrm{r} \times \mathrm{F}} \qquad \text{Eqn. 6-50}$$

The magnitude of the torque is the length of the lever arm times the component of the force normal to the wrench (Eqn. 6-51).

$$\|\boldsymbol{\tau}\| = rF\sin(\theta) \qquad \text{Eqn. 6-51}$$

Torque has units of Force times Distance or Joules (the same as Work or Energy).

Problem: The magnitude of the torque (or energy) needed to loosen a rusted bolt is 90 J. Calculate how much force you must apply with a wrench 10 cm = 0.1 m long.

Information: The unit of force is the Newton = 1 kg·m·s^{-2}. (This comes from, **F** = *m***a**, so although we tend to use kilograms for weight, that is only true when $g \approx 10$ m·s^{-2}.)

Solution: Assume you are smart enough to apply the force at right angles to the wrench so that sin(θ) = 1. Then the magnitude of the torque you need is,

$$\|\boldsymbol{\tau}\| = rF\sin(\theta) = rF \quad \Rightarrow \quad F = \frac{\|\boldsymbol{\tau}\|}{r} = \frac{90}{0.1} = 900\ \mathbf{N}$$

Interpretation: 900 Newtons ≈ 90 kilograms of weight on Earth since $g \approx 10$ m·s^{-2}. This means that if a person stood on the wrench he would have to weigh about 90×2.2 = 200 pounds to loosen the rusted bolt. If the wrench were 5 times longer you would only need to apply 1/5th the force (180 N or 40 pounds) and a child could do it.

Scenarios 2 and 3 are very strange. Imagine you are walking on a rotating platform such as a Merry-Go-Round (with no horses) that turns counterclockwise when you look down at it. As you walk you feel a force to your right, and the faster you walk the stronger the force gets. The same force occurs in the Northern Hemisphere on Earth; in the Southern Hemisphere it points to the left of the moving object. This so-called Coriolis Force is too small for us to notice, but winds and ocean currents feel it and are deflected by it, sometimes into fierce cyclones.

In Scenario 3 (Figure 6-25) imagine you are holding a Leyden Jar charged with protons. You walk to the East past a large magnet whose field points to the North. (By convention, magnetic fields point from the North to the South Magnetic Pole.) As you walk, the jar will be forced upward and it might levitate both itself and you if you walk fast enough.

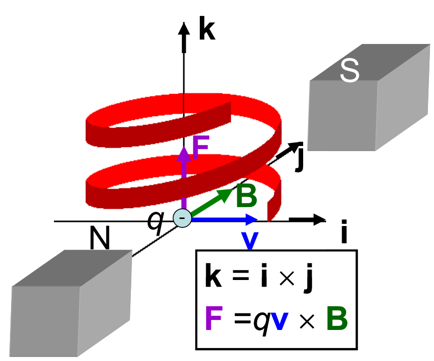


Figure 6-25 The Lorentz Force and cross product conform to the right hand screw rule.

Both these phenomena are described by the cross product and conform to the Right Hand Screw Rule illustrated in Figure 6-25. In the North Hemisphere, wind blows with low pressure on its left, not toward low pressure! When an electric charge, q that moves at velocity **v** crosses a magnetic field, **B** it experiences a force, **F** called the Lorentz force that acts at right angles to both the current and the magnetic field as in Figure 6-25, and given by the cross product,

$$\boxed{\mathbf{F} = q\mathbf{v} \times \mathbf{B}} \qquad \text{Eqn. 6-52}$$

Thus, when a proton moves in the positive x direction and the magnetic field points (from North to South Magnetic Poles) in the positive y direction, the proton is forced in the positive z direction. Similarly, when you turn a light bulb or a screw so that it moves into a socket or a board, it is designed so that as the fingers of your right hand rotate from the x to the y axis your extended thumb points and the bulb or screw moves in the positive z direction. (By the way, in the good old days, some incandescent light bulbs in the IND line of the NYC Subway were threaded in the opposite sense to avoid pilfering.)

When **v** and **B** are separated by an angle, θ the Lorentz force is,

$$F = qvB\sin(\theta) \qquad \text{Eqn. 6-53}$$

The Lorentz force is a maximum when **v** and **B** are oriented at right angles. It is zero when **v** and **B** are parallel. The moving charge must literally cross the magnetic field.

The cross product of vectors, **A** × **B**, was designed with these features in mind and includes the angle between the vectors. Thus, the magnitude of the cross product, ||**A** × **B**|| = $AB\sin(\theta)$ equals the area of a parallelogram with sides of length A and B separated by angle, θ. The cross product of two vectors = 0 when they are parallel and is maximum when they are perpendicular. The cross product is given by Eqn. 6-54 and its direction, **n**, follows the Right Hand Screw Rule, as in Figure 6-25. Note: Eqn. 6-54 includes a grid (highlighted in yellow) called a determinant.

$$\boxed{\begin{aligned} \mathrm{A}\times\mathrm{B} &= AB\sin(\theta)\mathrm{n} = \left(A_x\mathrm{i} + A_y\mathrm{j} + A_z\mathrm{k}\right)\times\left(B_x\mathrm{i} + B_y\mathrm{j} + B_z\mathrm{k}\right) = \\ &= \left(A_yB_z - A_zB_y\right)\mathrm{i} + \left(A_zB_x - A_xB_z\right)\mathrm{j} + \left(A_xB_y - A_yB_x\right)\mathrm{k} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \end{aligned}} \qquad \text{Eqn. 6-54}$$

The cross products of parallel unit vectors equal 0, namely, **i** × **i** = **j** × **j** = **k** × **k** = 0. The cross products of perpendicular unit vectors point in the 3rd perpendicular direction, according to the right hand screw rule, namely, **i** × **j** = **k**, **j** × **k** = **i** and **k** × **i** = **j**. Reversing the order of operations for the cross product reverses the sign and direction. Thus, for example, **i** × **j** = **k**, but **j** × **i** = **-k**.

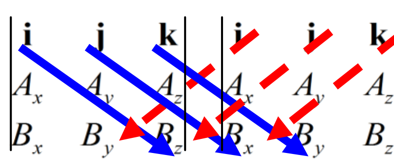


Figure 6-26 Graphical device for evaluating a Determinant.

The Determinant provides an easy way to calculate cross products if you follow the 3 steps listed below and illustrated in Figure 6-26.

1: Place two identical determinants side by side,
2: Add the products of terms along the blue lines,
3: Subtract the products of terms along the red lines.

Determinants also provide a simple way to arrange terms to solve systems of several equations with several unknowns.

Problem: Find the area of a parallelogram with sides whose vectors are, **A** = 2**i** + **j** +3**k** and **B** = 3**i** - **j** - **k**, and find the direction of the normal vector that is perpendicular to it.
Solution: Arrange the terms into repeated determinants and calculate using the 3 rules above. Then the cross product (which gives the direction of the normal vector, **n**) is,

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & 1 & 3 \\ 3 & -1 & -1 \end{vmatrix} \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & 1 & 3 \\ 3 & -1 & -1 \end{vmatrix} \Rightarrow \mathbf{n} = (-1+3)\mathbf{i} + (9+2)\mathbf{j} + (-2-3)\mathbf{k} = 2\mathbf{i} + 11\mathbf{j} - 5\mathbf{k}$$

Pythagoras gives the area of the parallelogram i. e, magnitude of the cross product as,

$$\mathbf{AREA} = AB\sin(\theta) = \sqrt{\left(2^2 + 11^2 + 5^2\right)} \approx 12.25$$

6. The Gradient

Now we get to Vector Calculus! The Gradient (∇f) is the vector that gives both the magnitude and direction of a slope or derivative of a function. It is given by Eqn. 6-55.

$$\boxed{\nabla f \equiv \mathbf{i}\frac{\partial f}{\partial x} + \mathbf{j}\frac{\partial f}{\partial y} + \mathbf{k}\frac{\partial f}{\partial z}} \qquad \text{Eqn. 6-55}$$

The gradient points in the direction that the function increases (= uphill). You can depict it with arrows pointing upslope or, from low to high concentrations of dots, as in Figure 6-27. The dots can represent any quantity, such as molecules, dust particles, or bugs. The length of the arrows must be proportional to the magnitude of the gradient. In Figure 6-27 the gradient (as well as the magnitude for these cases) of dot concentration is larger on the left side of the square and at the center of the circle.

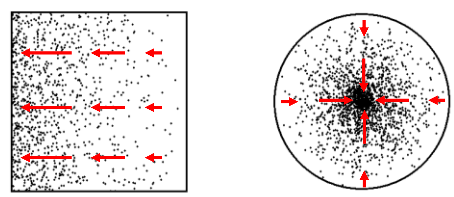


Figure 6-27. The gradient (red arrows) point from regions of low to high concentration of dots.

Problem: Find the gradient of Double Sine Mountain, from Eqn. 6-1 and Figure 6-2.
Solution: For this case, the function, f = z = 1.05·sin(x)·sin(y) from Eqn. 6-1 so we only need to take the first two terms on the right hand side of Eqn. 6-55. Thus,

$$\nabla z \equiv \mathbf{i}\frac{\partial z}{\partial x} + \mathbf{j}\frac{\partial z}{\partial y} = 1.05\left[\mathbf{i}\cos(x)\sin(y) + \mathbf{j}\sin(x)\cos(y)\right]$$

The slope or magnitude of Sine Mountain's gradient is given by Pythagoras as,

$$\left\|\nabla z\right\| \equiv 1.05\sqrt{\cos^2(x)\sin^2(y) + \sin^2(x)\cos^2(y)} = 1.05\left|\sin(x) - \sin(y)\right|$$

Over the domains, $0 \le x \le \pi$ and $0 \le y \le \pi$ both $\sin(x) \ge 0$ and $\sin(y) \ge 0$, so the maximum value of the slope is 1.05 and it occurs at the base of the mountain where either sin(x) = 1 and sin(y) = 0 or vice versa.

When a quantity moves in the direction of its gradient, expect it to change. For example, if lava from a volcanic eruption flows towards you, the height of the land will rise, and you had better run or you'll be buried. The rate of change of a quantity at any fixed point due to winds or currents or lava flows (called advection) equals minus the dot product of its gradient and the velocity, and is given by Eqn. 6-56.

$$\boxed{\frac{\partial f}{\partial t} = -\mathbf{v}\cdot\nabla f = -\left[v_x\frac{\partial f}{\partial x} + v_y\frac{\partial f}{\partial y} + v_z\frac{\partial f}{\partial z}\right]} \qquad \text{Eqn. 6-56}$$

When f = T, meteorologists use Eqn. 6-56 to forecast the rate of change of temperature due to the wind. This is useful when fronts pass by. Of course, accurate forecasts must also include solar heating, cooling at night, how the wind changes with time, etc.

7. Divergence

When you let air out of a tire the air diverges. Mass, density and pressure inside the tire decrease. Similarly, when a football game is over, the number of people in the stadium decreases as people diverge from it.

Divergence is the spreading rate of a vector. It is depicted in Figure 6-28 by dots that move or diffuse (ala Section 6-7) out of a circle with a porous perimeter.

Divergence of a vector, **V** is given by Eqn. 6-57,

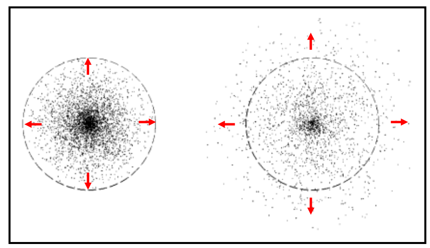


Figure 6-28. Sequence of dots diverging from a porous circle.

$$\boxed{\nabla\cdot\mathbf{V} = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_x}{\partial x}} \qquad \text{Eqn. 6-57}$$

The Fundamental Equation of Systems indicates that when Divergence is positive the quantity inside a volume will decrease with time unless there is a source creating it.

Problem: As a nova explodes, the velocity, **v** = 0.02(**i**x + **j**y + ***k***z) m·s^{-1} is radially outward. Calculate the value of the divergence at $r = 10^6$ km.
Hint: Since the flow is radially symmetric, you can choose any (x, y, z) values that satisfies ($x^2 + y^2 + z^2$) = r^2. Give yourself a break by choosing $y = z = 0$, so that $x = r$.
Solution: In that case $v_x = 0.02x$. Then, the divergence is,

$$\nabla\cdot\mathrm{v} = \frac{\partial v_x}{\partial x} = 0.02\ \mathrm{s}^{-1}$$

This means that the divergence or expansion is 2% per second, a considerable number.

8. **Curl and Vorticity**

The Curl, like the Cross Product, involves curious changes in direction. When an electric current, **J** runs along a straight wire it generates a magnetic field, **B** that curls around the wire, as in the left panel of Figure 6-29. The current is proportional to the Curl of the magnetic field. When air that rises in a severe thunderstorm has any initial rotation, the wind with velocity, **v** forms a rotating column or helix (recall Section 2.5) with vorticity, $\boldsymbol{\omega}$, that may intensity enough to form a tornado, as in the right panel of Figure 6-29. Vorticity is the curl of the velocity.

Curl, $\nabla \times \mathbf{V}$ is the rotation of a vector field. It is given by Eqn. 6-58.

$$\boxed{\begin{aligned}\nabla \times \mathbf{V} &= \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right) \times \left(V_x\mathbf{i} + V_y\mathbf{j} + V_z\mathbf{k}\right) = \\ &= \left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z}\right)\mathbf{i} + \left(\frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x}\right)\mathbf{j} + \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}\right)\mathbf{k}\end{aligned}} \qquad \text{Eqn. 6-58}$$

The curl, like the cross product is 1: governed by the right hand screw rule, and, 2: easily calculated by a determinant (Eqn. 6-59) using the approach of Figure 6-26.

$$\boxed{\nabla \times \mathbf{V} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ V_x & V_y & V_z \end{vmatrix}} \qquad \text{Eqn. 6-59}$$

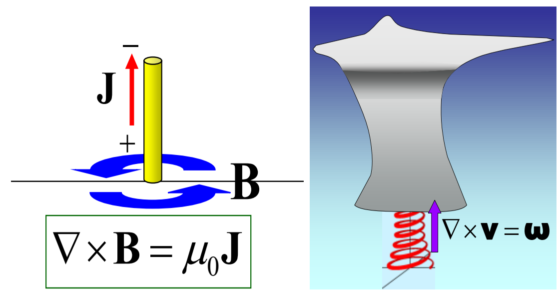


Figure 6-29. Left: An electric current generates a magnetic field with a curl. Right: Wind in severe thunderstorms often has a large curl or vorticity.

Applications of the curl lie at the foundation of the Industrial and Technological Revolutions. Without them the modern world would simply curl up and die. If you change a magnetic field by rotating a magnet in a wire loop, it will generate an electric current whose curl equals the rate of change of the magnetic field. This behavior is called Faraday's Law and it is the basis of the electric generator. Conversely, an electric current or a changing electric field generates a magnetic field with a curl. This behavior is called is Ampere's Law and it is the basis of the electromagnet and the electric motor.

Problem: Calculate the vorticity, $\boldsymbol{\omega}$ or the curl of a horizontal wind that blows in a counterclockwise sense around a low pressure area as a solid body with angular rotation rate, $\boldsymbol{\Omega} = \Omega_z\mathbf{k}$ around a vertical axis, where $\Omega_z = 10^{-4}\ s^{-1}$ (radians per second). With this rotation rate, the wind will complete a circle in period $\tau = 2\pi/(\Omega_z) \approx 6.28(10)^4$ s.
Information: Solid body rotation of horizontal motion is $\boldsymbol{\Omega} \times \mathbf{R}$, where $\mathbf{R} = x\mathbf{i} + y\mathbf{j}$.
Solution: First find velocity. The vector product for velocity is

$$\mathrm{v} = \boldsymbol{\Omega} \times \mathbf{R} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 0 & \Omega_z \\ x & y & 0 \end{vmatrix} = \Omega_z\left(-y\mathbf{i} + x\mathbf{j}\right)$$

Now, find the vorticity using Eqn. 6-56.

$$\boldsymbol{\omega} = \nabla \times \mathbf{v} = \Omega_z \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ -y & x & 0 \end{vmatrix} = \Omega_z\left(\frac{\partial x}{\partial x} + \frac{\partial y}{\partial y}\right)\mathbf{k} = 2\Omega_z\mathbf{k} = 2(10)^{-4}\mathbf{k}$$

Conclusion: Vorticity for solid body rotation is twice the angular rotation rate and is directed along the axis of rotation.

Don't Forget: Partial derivatives such as $\partial x/\partial z$ = 0 because x and z are independent variables and the partial derivative with respect to z holds both x and y constant.

Winds or currents that move in a straight line have curl or vorticity if the flow involves shear. Shear is a change of velocity at right angles to the flow. In any stream the current is faster away from the bank and up from the stream bed as in Figure 6-30. Any object immersed in shear flow will spin.

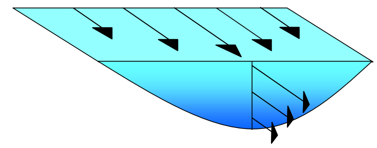


Figure 6-30. Stream flow with shear has vorticity.

9. The Laplacian: Poisson's Equation and Laplace's Equation

The Laplacian measures how concave or convex a function is. Concave or convex depends on your viewpoint, but note that the second syllable of conCave is Cave, and there is a tendency to fill a Cave (or a sock). I hope you remember from Section 2.8 that the second derivative is positive at a minimum. I really hope that you remember from this Chapter (Section 6.7) that when the second derivative of a function is positive, diffusion will cause it to increase with time, i. e, fill.

The Laplacian is the divergence of the gradient of a scalar function. In Cartesian Coordinates the 3-D version of the Laplacian is given by Eqn. 6-60.

$$\boxed{\nabla \cdot \nabla \phi \equiv \nabla^2 \phi = \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2}} \qquad \text{Eqn. 6-60}$$

The Laplacian is the common term in several of the most important partial differential equations, namely, the Classical Wave Equation (Eqn. 6-61: 1-D version, Eqn. 6-12), the Classical Diffusion Equation (Eqn. 6-62: 1-D version, Eqn. 6-21), and Poisson's Equation (Eqn. 6-63a) and Laplace's Equation (Eqn. 6-63b).

$$\nabla^2 \phi = \begin{cases} \dfrac{1}{c^2}\dfrac{\partial^2 \phi}{\partial t^2} & \text{Classical Wave Equation} \quad \text{Eqn. 6-61} \\ \dfrac{1}{D}\dfrac{\partial \phi}{\partial t} & \text{Classical Diffusion Equation} \quad \text{Eqn. 6-62} \\ 0 & \text{Laplace's Equation} \quad \text{Eqn. 6-63a} \\ f(x,y,z) & \text{Poisson's Equation} \quad \text{Eqn. 6-63b} \end{cases}$$

The profound differences between these equations and the situations they represent show up in the way time is represented. The second partial derivative with respect to time appears in the wave equation, where we have already seen that waves propagate out to infinity if allowed. Only the first partial derivative with respect to time appears in the diffusion equation, where any variations decrease with time. Both Laplace's and Poisson's Equations describe phenomena that have no explicit dependence on time such as the potential energy and forces on planets in the Solar System, stars in the Galaxy, and electric charges in a crystal (all neglecting Relativity), the stresses on a stretched membrane and steady flow of water through uniformly porous soils to rivers and wells.

The function, ϕ, takes on various different meanings depending on the equation. In the wave equation, ϕ can represent height of the surface for water waves, pressure for sound waves, or the electric field for electromagnetic waves. In the diffusion equation, ϕ can represent Temperature or concentration of some quantity. In both Poisson's and Laplace's Equations, ϕ typically represents a potential, such as potential energy.

Poisson's Equation and Laplace's Equation

Laplace's Equation is the homogeneous form of Poisson's Equation, i. e, when f(x, y, z) = 0. One way to state the difference is that Poisson's Equation describes the potential energy in a region with masses or electric charges while Laplace's Equation describes the potential energy when all masses or charges are outside the region.

Any function, ϕ that satisfies Laplace's Equation has several useful properties. **First**, once the value of the function is known on the outer boundary of a region, it can be solved at every point in the region. **Second**, both the maximum and minimum values of the function occur on the boundary of the region. **Third**, the function can vary but must do so in a way that is neither concave nor convex, as in Figure 6-31. One great consequence of the first two properties is that you cannot get electrocuted inside a conducting metal cage (a Faraday Cage) by lightning or live wires outside or touching the cage since the electric potential on the cage is constant, so it must also be constant inside the cage.

All this sounds great, but too often, Laplace's equation is too difficult to solve without the computer. So, what can we do? One thing we can do is work backwards from a solution, just as we did with the 1-D Classical Wave Equation, Eqn. 6-12. Similarly, the integral of many a function was found by working backwards, guessing the integral, then taking its derivative, and wishing, hoping, and praying we guessed right. If we weren't lucky we tried another. And if we got lucky we either found the function or found another function whose integral we had been trying to solve.

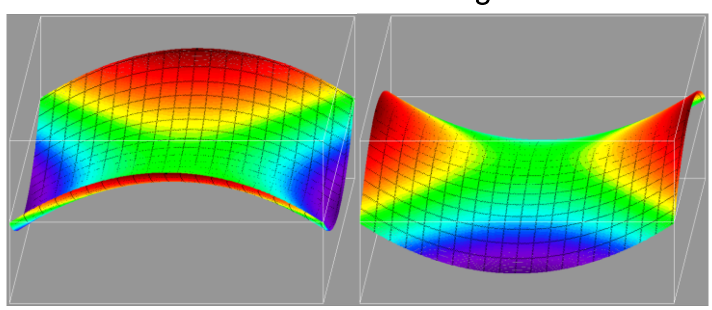


Figure 6-31. Graphs of $z = y^2 - x^2$ (left) and $z = x^2 - y^2$ (right) for $-1 \le x \le 1$ and $-1 \le y \le 1$. Both are saddle-shaped solutions of Laplace's Equation.

We will do the same for Laplace's Equation. It is actually easy to think up functions that satisfy it, and we can easily depict them if we restrict ourselves to 2-D problems because then $f(x, y)$ can be depicted as $z(x, y)$, the height of a surface. How about $z = ax + by$? It may not be particularly interesting - it is simply a tilted plane - but it satisfies Laplace's Equation because all second partial derivatives are 0. The function, $z = xy$ is a bit more interesting because it is a curved surface. So too are $z = \pm(x^2 - y^2)$, which are particularly interesting because they are the equations of saddles, as in Figure 6-31.

Saddled with this introduction to Vectors and Vector Calculus, we take our chances to ride into the sunset.

6.14 Chances Are: Statistics and The Normal Distribution

At long last you have reached the grand finale. All through the book you have learned that when life or math throws you a curve, Calculus will find its area, its rate of change, its past, and its future. But I have saved the best face of Calculus for last. Calculus is the premier way to handle the inevitable uncertainties in life. If it weren't, why would the banks, insurance companies and investment houses hire armies of actuaries and statisticians?

In this section we examine the wondrous function, e^{-kx^2} (called the Gaussian, after you know who) and its integral, Eqn. 6-64 because they are of central value in statistics. The integral is a probability or frequency distribution of a quantity, x, that can be molecular speed or distance from a target in a specific direction, or deviation from the mean of human height, longevity, IQ or test scores.

$$\boxed{P(x_1 \le x \le x_2) = \sqrt{\frac{k}{\pi}}\int_{x_1}^{x_2} e^{-kx^2}\,\mathrm{d}x} \qquad \text{Eqn. 6-64}$$

The constant, $[k/\pi]^{1/2}$ outside the integral of Eqn. 6-64 is there to ensure (as we prove below) that total probability over the entire domain, $-\infty \le x \le \infty$ is $P = 1$ (100%).

Eqn. 6-64 also belongs here for a purely mathematical reason. You may remember that it is equivalent to Eqn. 3-36, one of the 'impossible' integrals of Section 3-12, but with a more general domain. There is only one case that this integral can be solved analytically, namely when the domain is, $-\infty \le x \le +\infty$ (or by symmetry, $0 \le x \le +\infty$), and then we need to use three tricks in a row. Tricks #1 and #2 set up Trick #3. Remember also that for Eqn. 6-64 to represent a probability, we must prove the constant outside the integral = $[k/\pi]^{1/2}$, for then the total probability or integral over the entire domain, $P(-\infty \le x \le \infty) = 1$.

Trick #1. Square the integral in Eqn. 6-64. This is the first step of genius.
Trick #2. Change one of the variables from x to y since the variable name doesn't matter.

$$\left[\int_{-\infty}^{\infty} e^{-k(x^2)}\mathrm{d}x\right]\left[\int_{-\infty}^{\infty} e^{-k(y^2)}\mathrm{d}y\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-k(x^2+y^2)}\mathrm{d}x\mathrm{d}y$$

Trick #3. Transform the Double Integral to Cylindrical Coordinates, noting that $x^2 + y^2 = R^2$. Furthermore, since $R\mathrm{d}R = ½\mathrm{d}(R^2)$, the integrand is a perfect differential. Integrating θ from 0 to 2π produces a factor of 2π.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-k(x^2+y^2)}\mathrm{d}x\mathrm{d}y = \int_0^{\infty}\int_0^{2\pi} R\cdot e^{-kR^2}\,\mathrm{d}\theta\mathrm{d}R = \frac{1}{2k}\int_0^{\infty}\int_0^{2\pi} e^{-kR^2}\,\mathrm{d}\theta\mathrm{d}(kR^2) = \frac{\pi}{k}\int_0^{\infty} e^{-kR^2}\,\mathrm{d}(kR^2) = \frac{\pi}{k}$$

Take the square root to restore the initial integral

$$\sqrt{\frac{\pi}{k}} = \int_{-\infty}^{\infty} e^{-kx^2}\,\mathrm{d}x \quad \Leftrightarrow \quad 1 = \sqrt{\frac{k}{\pi}}\int_{-\infty}^{\infty} e^{-kx^2}\,\mathrm{d}x$$

This proves that Eqn. 6-64 represents a probability for any interval.

Figure 6-32 shows narrow and wide versions of the Gaussian function, e^{-x^2}. Do these curves ring a bell? Think about it before continuing to read....They should ring a bell because they are versions of the Normal Distribution, better known as the Bell Curve. In populations that obey the Normal Distribution individuals near the mean are most numerous and the further from the mean, the fewer individuals.

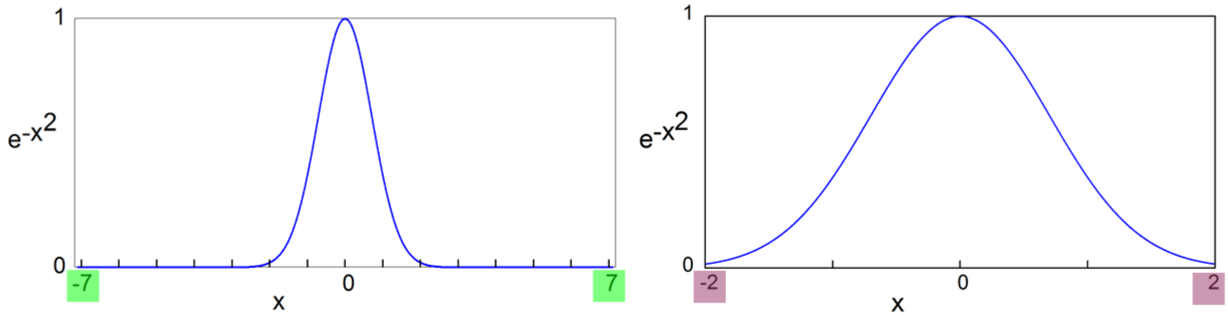


Figure 6-32 Graphs of the Gaussian, e^{-x^2}. The left graph fits Fig. 1-16 and extends from $-7 \le x \le 7$. The right graph is stretched sideways and so, it extends only from $-2 \le x \le 2$.

The mean values of the curves in Figure 6-32 are 0 because e^{-x^2} is symmetrical about 0. If the mean value of a quantity (e. g., human height) is $\overline{x}$, simply move the curve to the right by $\overline{x}$ and use $(x - \overline{x})^2$ in place of x^2 in the integrand of Eqn. 6-64. The constant, k in the Gaussian is inversely proportional to the spread of the curve.

Note: Even though the Normal Distribution extends from $-\infty \to +\infty$, it accurately represents many quantities, such as human height or test grades that are always positive and have finite limits. Simply ignore the tiny incorrect probabilities given by the Normal Distribution for unrealistic or impossible heights, weights, or grades.

Statistical Measures: Means, Variances and Standard Deviations

The mean of a population is the first important general property of all statistical distributions. The second important general property is the variability. The standard measure of variation of a population is called the Standard Deviation.

Means and weighted means of continuous functions were defined in Section 3.10 and we will return to them presently. But first we consider functions that consist of discrete series of numbers. The mean value, $\overline{x}$ of a discrete series of numbers x_1, x_2, x_3, $\cdots$ x_n, is the arithmetic average, or the sum of the numbers divided by the number of terms.

$$\boxed{\overline{x} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i} \qquad \text{Eqn. 6-65}$$

Any number, x_i deviates from the average $\overline{x}$ by an amount, x'.

$$\boxed{x_i \equiv \overline{x} + x_i{}'} \qquad \text{Eqn. 6-66}$$

Problem: A person's IQ is 85 and average IQ is $\overline{x} = 100$. Find the deviation from average
Solution: Using Eqn. 6-66, $x = \overline{x} + x'$ = 85 = 100 + x'. Solving for x' yields x' = -15.

The Standard Deviation, σ, is a measure of the typical size of variations of individuals in a population or of numbers in a series. To calculate it,

1: Square the deviation from the mean of each of *n* individuals or numbers,
2: Add all the squared deviations
3: Divide by (*n* - 1). This is the variance, σ^2, the square of the standard deviation.
4. Take the square root.

$$\boxed{\sigma \equiv \sqrt{\overline{x'^2}} \equiv \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n} x_i'^2}} \qquad \text{Eqn. 6-67}$$

We square deviations to get a positive sum; by definition, the sum of deviations = 0. The reason to divide by (*n* - 1) instead of *n* in Eqn. 6-67 is that the number of independent deviations (called degrees of freedom) is 1 less than *n*. If a sample consists of *n* = 1 term there are 0 variations. If a sample has *n* = 2 terms, and the first is 5 above average, the second is compelled to be 5 below average (so it is not independent).

Eqn. 6-67 may create confusion until we do an example. Let's analyze the heights of 5 adult males given in column 2 of Table 6-5. (I chose these heights to approximate the actual mean and standard deviation of the world's men.)

Name	Height	x'	x'^2
x_1	66	-3	9
x_2	73	4	16
x_3	68	-1	1
x_4	67	-2	4
x_5	71	2	4
$\overline{x}$	69	σ	2.915

Table 6-5 Heights, deviations, x' and squares of deviations of 5 men with mean and standard deviation.

Problem: Calculate the Standard Deviation of the Men's heights in Table 6-5.
Solution: Plug values of x' from Table 6-5 into Eqn. 6-74.

$$\sigma = \sqrt{\frac{1}{(5-1)}\left[(-3)^2 + 4^2 + (-1)^2 + (-2)^2 + 2^2\right]} = \sqrt{\frac{1}{4}[9+16+1+4+4]} \approx 2.915$$

Variance and Standard Deviation for the Binomial Distribution

The general equations for mean and standard deviation of the *binomial* distribution are,

$$\boxed{\overline{x}_{binom} = np} \qquad \text{Eqn. 6-68}$$

$$\boxed{\sigma_{binom} \equiv \sqrt{np(1-p)}} \qquad \text{Eqn. 6-69}$$

Problem: Calculate the Mean and Standard Deviation of the binomial distribution for n = 100 and p = 0.5
Solution: Plug into Eqn. 6-68 and Eqn. 6-69.

$$\bar{x}_{binom} = np = 100 \times 0.5 = 50$$

$$\sigma_{binom} \equiv \sqrt{np(1-p)} = \sqrt{100 \times 0.5(1-0.5)} = 5$$

Variance and Standard Deviation for the Normal Distribution

For any continuous function or distribution, the standard deviation is an integral. Perhaps you remember way back in Section 3.10 the treatment of means and weighted means. The mean value of a distribution is actually a weighted mean, where the quantity *x* (e. g., height, IQ, salary) is the weighting function. The Normal Distribution's mean is,

$$\boxed{\bar{x} = \sqrt{\frac{k}{\pi}} \int_{-\infty}^{\infty} x e^{-kx^2} \mathrm{d}x = 0}$$ Eqn. 6-70

You can either do this integral directly noting that $x\mathrm{d}x = \mathrm{d}(x^2/2)$ or by noting, as I pointed out a few paragraphs ago, that the Gaussian is symmetric about 0. Multiplying it by x makes it antisymmetric about $x = 0$ so the area under the curve, $x e^{-x^2} = 0$.

The mean value of the Gaussian is called its first moment because it is multiplied by x^1. The Variance is a weighted mean using the second moment of the Gaussian because it is multiplied or weighted by x^2. (Through the irony of math, they are called moments even though they take can hours to calculate.)

$$\boxed{\overline{x'^2} \equiv \sigma^2 = \frac{\int_{-\infty}^{\infty} x^2 e^{-kx^2} \mathrm{d}x}{\int_{-\infty}^{\infty} e^{-kx^2} \mathrm{d}x} = \sqrt{\frac{\pi}{k}} \int_{-\infty}^{\infty} x^2 e^{-kx^2} \mathrm{d}x}$$ Eqn. 6-71

The integrand on the right of Eqn. 6-71 involves the product, $x^2 e^{-kx^2}$. That calls for integration by parts. Warning: Choosing u and v is tricky! Choosing $u = x^2$ or $u = e^{-kx^2}$ leads up a blind alley. Choosing $u = x$ [and therefore $\mathrm{d}v = x e^{-kx^2} \Rightarrow v = -e^{-kx^2}/(2k)$] works perfectly. Doing the integration by parts is then a piece of cake.

$$\sqrt{\frac{\pi}{k}} \int_{-\infty}^{\infty} x^2 e^{-kx^2} \mathrm{d}x = -\frac{1}{2k}\sqrt{\frac{\pi}{k}}\left[x e^{-kx^2} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-kx^2} \mathrm{d}x \right] = -\frac{1}{2k}\sqrt{\frac{\pi}{k}}\left[0 - \sqrt{\frac{k}{\pi}}\right]$$

The radicals cancel, leaving the equation for Variance, σ^2 as,

$$\boxed{\sigma^2{}_{normal} = \frac{1}{2k}}$$ Eqn. 6-72

We now make two changes to Eqn. 6-64. First, use Eqn. 6-72 to replace k with $1/(2\sigma^2)$. Second, generalize to allow quantities with mean value, $\bar{x} \neq 0$. This leads to the classic equation for

The Normal Distribution

$$P(x_1 \le x \le x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \mathrm{d}x \qquad \text{Eqn. 6-73}$$

Eqn. 6-73 gives the probability, P that a variable, x (such as human height) lies between x_1 and x_2. Table 6-6 gives vital information and insight about the normal distribution. Let's apply this to the heights of adult American men. The mean ≈ 69″ (inches) and the standard deviation, $\sigma \approx 2.9''$. The second column of Table 6-6 means that ≈ 16% of men are taller than 1σ above the mean height or 69 + 2.9 = 71.9″, while another 16% are shorter than 1σ below the mean height or 69 - 2.9 = 66.1″. An even smaller percentage, namely only ≈ 2.3% are taller than 2σ above the mean or 69 + 2×2.9 = 74.8″ and another 2.3% are shorter than 69 - 2×2.9 = 63.2″. A much smaller percentage of men ≈ 0.135% are taller than 3σ above the mean or shorter than 3σ below the mean.

# of st dev from mean	% for normal distb	% for binomial distb
0	50	50
1	15.866	16.0
2	2.275	2.302
3	0.135	0.133
4	0.0032	0.0028
5	3E-05	1.86E-05

Table 6-6 Frequencies for the Normal Distb. and the Binomial Distb. with n = 100.

So it goes for many characteristics. Take IQ! Average IQ is set to 100. The standard deviation of IQ's is about 15, so that only 16% of people have IQ's above 115. If you have understood everything in this book then your IQ is probably at least 3 standard deviations above the mean or 100 + 3×15 = 145. Of course in terms of personality and looks, you are all at least 5 standard deviations above the mean, or 1 out of 3.3 million!

Testing Honesty: The Null Hypothesis

Probability and statistics provide us with an invaluable technique to test honesty. The smaller the probability of an event and the further it is from the mean, the less likely it occurred by chance or indeed, the less likely it occurred at all. This is called the Null Hypothesis. Thus, for example, if someone tells you he flipped a coin 100 times and got 80 heads you can be pretty confident that the coin is biased, or he either cheated or is lying. If someone tells you that his abilities are 6 standard deviations above the mean it is more likely that his degree of honesty is several standard deviations below the mean. The moral: If you know probability and statistics it won't be easy to trick you.

For Large Numbers the Binomial Distribution Approaches the Normal Distribution

Figure 6-33 repeats Figure 1-15 in showing the Binomial Distribution for n = 4, 10, 20, and 100, but also compares it to the Normal Distribution and how that match improves as n increases. Table 6-6 compares the frequencies of the Normal Distribution and Binomial Distribution for flipping coins n = 100 times. The match is amazingly close when σ < 4. The two distributions do deviate from each other for the extremely rare cases when σ > 5. This is because the Normal Distribution has no limit but the Binomial Distribution does - you can't get more than 100 heads when you toss a coin 100 times.

I find it amazing that the distribution of human heights, IQ's, and longevity all conform to the Normal Distribution you get when you toss a large number of coins many times. You can feel anything you want about this amazing coincidence, which seems to imply that our IQ's, heights, test scores are all governed by a random process akin to flipping coins, but one sure conclusion is that genetic variations are governed by a similar statistical principle that governs the distribution of all random quantities. The mathematical tendency for large samples of random variables to approach the Normal Distribution is called the Central Limit Theorem.

Proving that the visually obvious match between the Binomial and Normal Distributions (as $n \to \infty$) is true mathematically, is a far trickier matter. The proof hinges on showing that for large numbers, the factorial functions of the Binomial Distribution merge with the exponential function of the Normal Distribution. I will not show the complete proof, which you can look up in a statistics book. All I do here is to start the proof by relating factorials to exponentials.

The first step is to take the natural log of $n!$ and remember that the log of a product of terms is the sum of the logs of the terms

$$\ln(n!) = \ln(1) + \ln(2) + \ln(3) + \ldots + \ln(n)$$

Next, use Eqn. 3-20 (Did you remember that?) to integrate ln(x) from 1 to n.

$$\boxed{\int_1^n \ln(x)\mathrm{d}x = n\ln(n) - n + 1} \qquad \text{Eqn. 6-74}$$

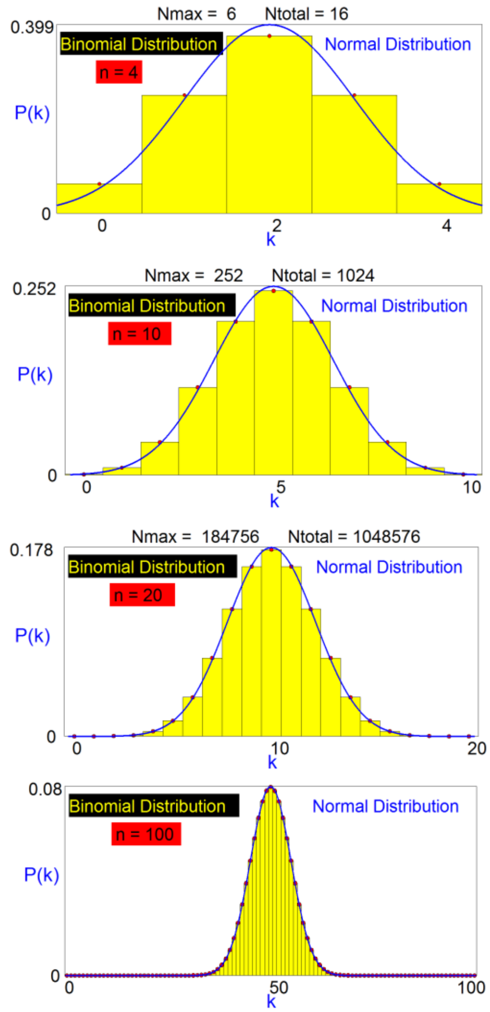


Figure 6-33 Normal vs Binomial distribution for n = 4, 10, 20, 100.

Eqn. 6-74 is slightly larger than the left hand Riemann sum of boxes with $\Delta x = 1$.

$$\boxed{\int_1^n \ln(x)\mathrm{d}x >\approx \ln(1) + \ln(2) + \ln(3) + \ldots + \ln(n) = \ln(n!)} \qquad \text{Eqn. 6-75}$$

Combining Eqn. 6-74 and Eqn. 6-75 yields

$$\boxed{\ln(n!) >\approx n\ln(n) - n + 1 <\approx \ln\left(n^n\right) - n} \qquad \text{Eqn. 6-76}$$

The antilog of Eqn. 6-76 is the result we want because it shows that for large numbers, the factorial is indeed related to an exponential.

$$\boxed{n! \approx n^n e^{-n}} \qquad \text{Eqn. 6-77}$$

Eqn. 6-77 was derived by James Stirling in 1768. A sterling (better) approximation is,

$$\boxed{n! \approx \sqrt{2\pi n}\, n^{n+0.5} e^{-n}} \qquad \text{Eqn. 6-78}$$

Problem: Test the accuracy of Eqn. 6-78 when n = 5 and when n = 8.
Solution:

$$5! = 120 \approx \sqrt{2\pi}\, 5^{5.5} e^{-5} \approx 118.0$$

$$8! = 40320 \approx \sqrt{2\pi}\, 8^{8.5} e^{-8} \approx 39902$$

The percentage errors are 1.65% and 1.04% respectively, and continue to decrease as n increases. For example, when n = 50 the percentage error of Eqn. 6-78 is 0.167%.

So, we have shown that, at least for large numbers, the Binomial Distribution acts like the Normal Distribution. The random variations along the long helices of DNA molecules play a statistical game that has evolved us into geniuses. All is well with math and with the world. Thus, have we finished our long voyage through the Calculus.

6.15 Farewell and I'll be Seeing You!

We have arrived at the end. You should revel in the fact that you have undertaken one of humankind's great intellectual odysseys, an adventure that the Modern World has hinged on. The Calculus has been an integral (pun intended) part of the scientific and technological revolutions that have extended human life span and opened so many marvelous opportunities and joys to us. By dint of persistence and hard work, you have traversed your Royal Road to Genius. So, my new geniuses, and if I have the honor, my friends, you have every right to feel proud of your accomplishment and to celebrate your triumph.

And the best way to celebrate that triumph is to take a victory lap. So, without delay, if this was your first time reading this book, start reading it over again from beginning to end, and once you are back here you may really have just mastered Calculus!

Index