Math 110, Fall 2017, Course Packet

Contents

IV Multivariable calculus 80

Part I

Using the math you already know

1 Functions

If we count AB calculus as a pre-requisite and pre-calculus/trigonometry is a pre-pre-requisite, then functions and their graphs are a pre-pre-pre-requisite! But... that doesn't mean that most of you are sufficiently good at dealing with these. Recognition of basic types of functions is crucial for being able to handle material at the pace and level you will need. So is the ability to go back and forth between analytic expressions for functions and their graphs. So is number sense: knowing approximate values without stopping for a detailed calculuation.

Because of the preliminary nature of this material, I am not going to write comprehensive notes on it. Instead, I will assign some online homework to make sure you are where you need to be and will refer you to sections of the textbook to brush up. Here are the key concepts and vocabulary from Section 1.1 of the textbook. Know these!

- domain
- range
- notation for piecewise defintions
- absolute value function
- greatest integer function
- increasing function
- decreasing function
- even function
- odd function

Suggested reading if any of this is unfamilair: Sections 1.3 and 1.6 of Hughes-Hallett et al., *Calculus, 5th edition*.

If *too* much of this is unfamiliar, you may be in the wrong course!

1.1 Graphing

Begin by reading Section 1.2 of Thomas, paying particular attention to the following topics: composition of functions; shifting a graph; scaling a graph; reflecting a graph and reflectional symmetry. Also please skim Sections 1.3 and 1.4, though we will not be emphasizing these.

Tips on graphing an unfamiliar function, f

(i) Is the domain all real numbers, or if not, what is it? If the function has a piecewise definition, try drawing each piece separately.

(ii) Is there an obvious symmetry? If $f(-x) = f(x)$ then f is even and there is a symmetry about the y-axis. If $f(-x) = -f(x)$ then f is odd and there is 180-degree rotational symmetry about the origin.

(iii) Are there discontinuities, and if so where? Are there asymptotes?

(iv) Try values of the function near the discontinuities to get an idea of the shape – these are particularly important places.

(v) Try computing some easy points. Often $f(0)$ or $f(1)$ is easy to compute. Trig functions are easily evaluated at certain multiples of π.

(vi) Where is f positive? Where is f increasing (where is $f' > 0$)? Where is f concave upward ($f'' > 0$), versus downward ($f'' < 0$)?

(vii) Where are the maxima and minima of f and what are its values there?

(viii) Is f defined everywhere? If not, what is the domain?

(ix) What does f do as $x \to \infty$ or $x \to -\infty$?

(x) Is there a function you understand better than f which is close enough to f that their graphs look similar?

(xi) Is f periodic? Most combinations of trig functions will be periodic.

1.2 Proportionality, units and applications

Another skill most students need practice with is writing formulas for functions given by verbal descriptions. For example, knowing that an inch is 2.54 centimeters, if $f(x)$ is the mass of a bug x centimeters long, what function represents the mass of a bug x inches long?

(a) $2.54f(x)$

(b) $f(x)/2.54$

(c) $f(2.54x)$

(d) $f(x/2.54)$

It helps to think about all such problems in units. Although inches are bigger than centimeters by a factor of 2.54, numbers giving lengths in inches are *less than* numbers giving lengths in centimeters by exactly this same factor. Writing this in units prevents you from making a mistake:

$$x \text{ cm} \times \frac{1 \text{ in}}{2.54 \text{ cm}} = \frac{x}{2.54} \text{ in}\,.$$

This shows that replacing x by $x/2.54$ converts the measurement, and therefore (d) is the correct answer. OK, maybe that was too easy for you, but when the problems get more complicated, it really helps to do this.

Some more helpful facts about units are as follows.

1. You can't add or subtract quantities unless they have the same units.

2. Multiplying (resp. dividing) quantities multiplies (resp. divides) the units.

3. Taking a power raises the units to that power. Most functions other than powers require unitless quantites for their input. For example, in a formula $y = e^{***}$ the quantity *** must be unitless. The same is true of logarithms and trig functions: their arguments (inputs) are always unitless.

4. Units tell you how a quantity transforms under scale changes. For example a square inch is $2,54^2$ times as big as a square centimeter, and a Netwon (kilogram meter per second squared) is 10^5 dynes (gram centimeter per second squared).

Often what we can easily tell about a function is that it is proportional to some combination of other quantities, where the **constant of proportionality** may or may not be known, or may vary from one version of the problem to another.

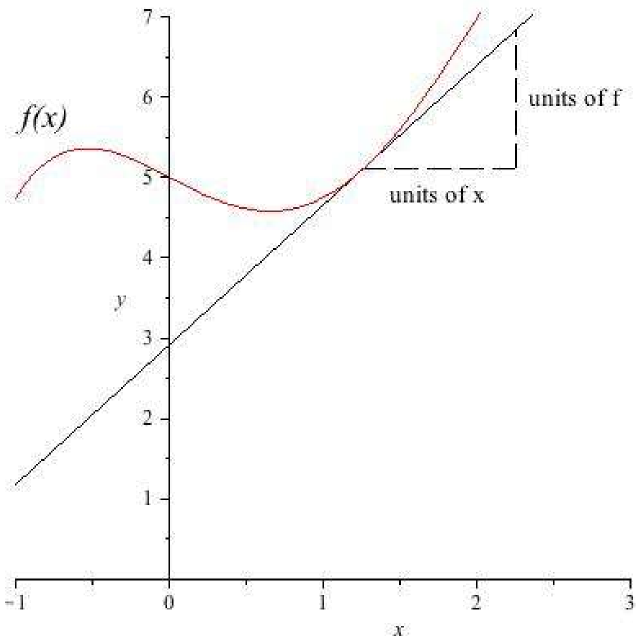
EXAMPLE: if the monetization of a social networking app is proportional to the square of the number of subscribers (this representing perhaps the amount of messaging going on) then one might write $M = kN^2$ where M is monetization, N is number of subscribers and k is the constant of proportionality. You should always give units for such constants. They can be deduced from the units of everything else. The units of N are people and the units of M are dollars, so k is in dollars per square person. You can write it: k \$/person^2.

EXAMPLE: The present value under constant discounting is given by $V(t) = V_0 e^{-\alpha t}$ where V_0 is the initial value and α is the discount rate. What are the units of α? They have to be inverse time units because αt must be unitless. A typical discount rate is 2% per year. You could say that as "0.02 inverse years."

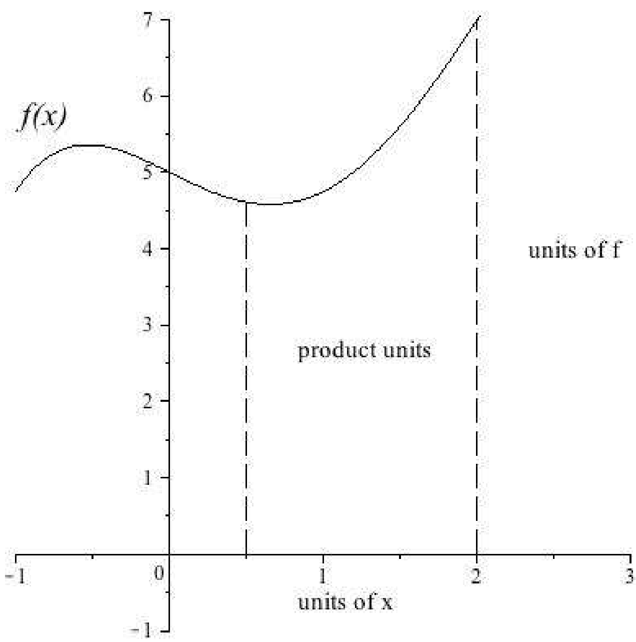
Often quantities are measured as proportions. For example, the proportional increase in sales is the change is sales divided by sales. In an equation: the proportional increase in S is $\Delta S/S$. Here, ΔS is the difference between the new and old values of S. You can subtract because both have the same units (sales), so ΔS has units of sales as well. That makes the proportional increase unitless. In fact proportions are always unitless.

Percentage increases are always unitless. In fact they are proportional increases multiplied by 100. Thus if the proportional increase is 0.0183, the percentage increase is 18.3%. In this class we aren't going to be picky about proportion versus percentage. If you say the percentage increase is 0.183 or the proportional change is 18.3%, everyone will know exactly what you mean. But you may as well be precise.

One last thing about units (really should have been point number 5 above) is how they behave under differentiation and integration. The derivative $(d/dx)f$ has units of f divided by units of x. You can see this easily on the graph because it's a limit of rise over run where rise has units of f and run has units of x. Likewise, $\int f(x)\,dx$ has units of f times units of x. Again you can see it from the picture, because the integral is an area under a graph where the y-axis has units of f and the x-axis has units of, well, x.



f(x)
units of f
units of x
y
x



f(x)
units of f
product units
units of x

Inverse functions

You can read about inverse functions is Section 1.6 of the textbook. The concept appears to be harder than most people realize. For example, last year I gave a problem to compute a formula for the inverse function to $\sinh(x)$, the hyperbolic sine function. This was already not easy (only half the students got it) but then I asked them to find a number u such that $\sinh(u) = 1$. Almost no one got it, despite that fact that this was supposed to be the easy part! They just needed to plug in 1 to their inverse sinh function. By definition, $\sinh^{-1}(1)$ is a number u such that $\sinh(u) = 1$. The moral of the story is, don't lose sight of the meaning of an inverse function when doing computations with them.

The textbook tells you how to compute them and also some things to watch out for:

- When does an inverse function exist?
- What are the domain and range of the inverse function?

When a function is not one to one, you can define an inverse if you restrict the range of the inverse. There is a standard way this is done with inverse trig functions; please read it on page 48. Note: the standard inverse trig functions have names (arcsine, etc.) and notations ($\sin^{-1}$ and so forth). Note also, that the notation f^{-1} for the inverse function to f is TERRIBLE. It is the same as (and therefore confusable with) the notation of the reciprocal of f. How stupid is that? But it's widely accepted, so we're stuck with it. Another example of a standard choice of inverse function is the inverse of the squaring function.

> Think: how is the squaring funtion not one to one, what is the name of its standard inverse, and what choice is made to remedy its lack of being one-to-one?

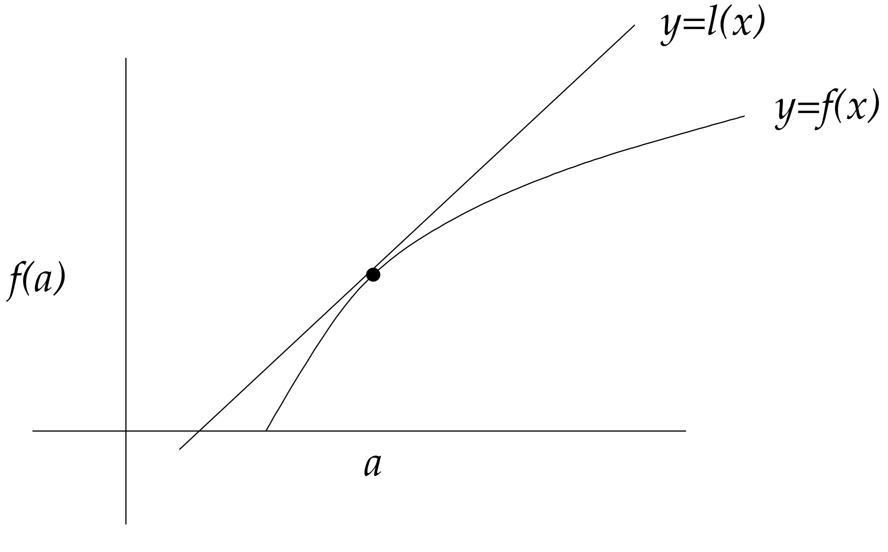
Lastly, think about the units of an inverse function. If f takes units of x as input and produces units of y, then to answer the question "f of what is equal to y?" you need to input a quantity in units of y and answer in units of x. In other words, the input and output units are switched.

1.3 Estimating and bounding

Estimating is non-rigorous. We want to understand a quantity $f(x)$ that is hard to compute, but we can compute a quantity $\tilde{f}(x)$ that is near $f(x)$. It's a little subjective to say $\tilde{f}(x) \approx f(x)$, if we can't say precisely what is meant by the symbol $\approx$, but it is very useful nonetheless. One of the most common methods of approximation is the linear approximation via the derivative. This is discussed at length in Section 3.11, too much length actually. The introduce the term *differential* which we won't use. We will however use the term *linearization*:

> **Definition:** The linearization of a differentiable function f at the point a is the function $\ell(x) = f(a) + (x - a)f'(a)$ (see page 203 of the text). In pictures $\ell(x)$ is the function whose graph is a line tangent to the graph of f at the point $(a, f(a))$.

How close is $\ell(x)$ to $f(x)$ when x is close to a? Obviously it depends how close x is to a. In the middle of the course, when we study Taylor polynomials, we'll see that $\ell(x) - f(x)$ is roughly a constant times $(x - a)^2$. Squaring makes small numbers even smaller, so when x is within 0.01 of a, then $\ell(x)$ should be within a few ten-thousandths of $f(a)$. We can't get more precise than this at present, but it's good to keep in mind.



Bounding

Bounding is rigorous. To get an upper bound on $f(x)$ means to find a quantity $U(x)$ that you understand better than $f(x)$ for which you can prove that $U(x) \geq f(x)$. A lower bound is a quantity $L(x)$ that you understand better than $f(x)$ and that you can prove to satisfy $L(x) \leq f(x)$. If you have both a lower and upper bound, then $f(x)$ is stuck for certain in the interval $[L(x), U(x)]$. It should be obvious that an upper bound is better the smaller it is. Similarly, a lower bound is better the larger it is.

In a way, though, bounding is harder than estimating because there is no one correct bound (well there's no one correct estimate either, but we usually a particular estimate we're told to use, such as a linear estimate). Two ways we typically find bounds are as follows.

First, if f is monotone increasing then an easy upper bound for $f(x)$ is $f(u)$ for any $u \geq x$ for which we can compute $f(u)$. Similarly an easy lower bound is $f(v)$ for any $v \leq x$ for which we can compute $f(v)$. If f is monotone decreasing, you can swap the roles of u and v in finding upper and lower bounds. There are even stupider useful bounds, such as $f(x) \leq C$ if f is a function that never gets above C.

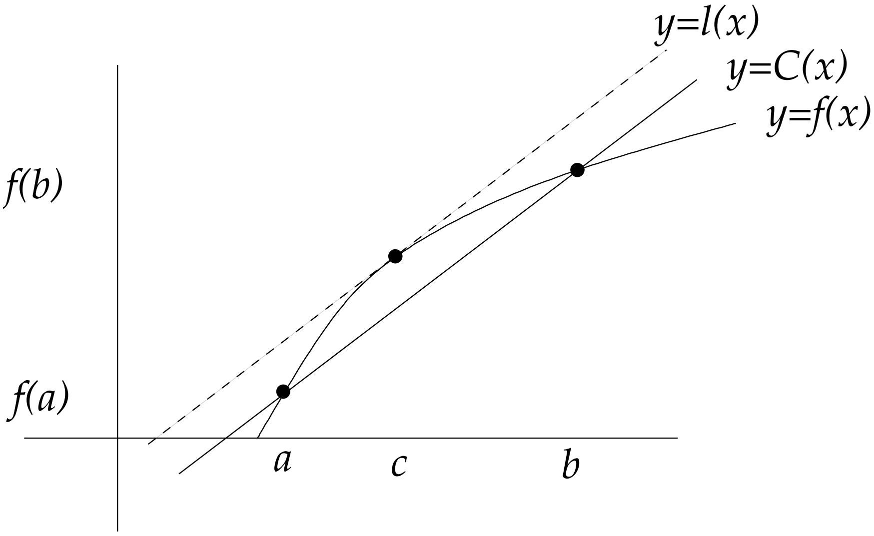
Example: Suppose $f(x) = \sin(1)$. The easiest upper and lower bounds are 1 and -1 respectively because sin never goes aboe 1 or below -1. A better lower bound is 0 because $\sin(x)$ remains positive until $x = \pi/2$ and obviously $1 < \pi/2$. You might in fact recall that one radian is just a bit under 60°, meaning that $\sin(60°) = \sqrt{3}/2 \approx 0.0866\ldots$ is an upper bound for $\sin(1)$. Computing more carefully, we find that a radian is also less than 58°. Is $\sin(58°)$ a better upper bound? Probably not because we don't know how to calculate it, so it's not a quantity we understand better. Of course is we had an old-fashioned table of sines, and all we can remember about one radian is that it is between 57° and 58°, then $\sin(58°)$ is not only an upper bound but the best one we have.

Concavity

A more subtle bound come when f is know to be concave upward or downward in some region. By definition, a concave upward function lies below its chords and a concave downward function lies above its chords.

Concavity upward (resp. downward) is easy to test: a function f is concave upward wherever $f'' > 0$ and concave downward wherever $f'' < 0$. If one of these holds over an interval $[a, b]$ then for x in the interval $[a, b]$ you can tell whether $f(x)$ is greater or less than the chord approximation

$$C(x) = f(a) + \frac{x-a}{b-a}(f(b) - f(a)).$$



In the figure, the function $f(x)$ is concave down. as long as x is in the interval $[a, b]$, we are guaranteed to have $C(x) \leq f(x)$. On the other hand, when $f'' < 0$ on an interval, the function always lies below the tangent line. Therefore $L(x)$ is an upper bound for $f(x)$ when $x \in [a, b]$ no matter which point $c \in [a, b]$ at which we choose to take the linear approximation.

Example: The function $\tan x$ is concave upward on $[0, \pi/2)$. That means that the tangent line to $\tan x$ anywhere in that interval will be a lower bound for $\tan x$ on the interval. The easiest place to compute the slope of $\tan x$ is at $x = 0$, where thederivative is $\sec^2(0) = 1$. The tangent line at $(0, 0)$ is therefore $y = x$. This gives the lower bound $\tan x \geq x$. This is in fact VERY close when x is near zero because tan has a point of inflection there (the tangent line pases through the curve, which is particularly flat).

1.4 Limits

You might not think limits would show up in a calculus course oriented toward application. Wrong! There are a lot of reasons why you need to understand the basic of limits. You should know these reasons, so here they are.

1. The definition of derivative (instantaneous rate of change) is a limit.
2. The number e is defined by a limit.
3. Continuous compounding is a limit.
4. Limits are needed to understand improper integrals, such as the integrals of probability densities.
5. Infinite series, which we will discuss briefly, require limits.
6. Discussing relative sizes of functions is really about limits.

I would like you to understand limits in four ways:

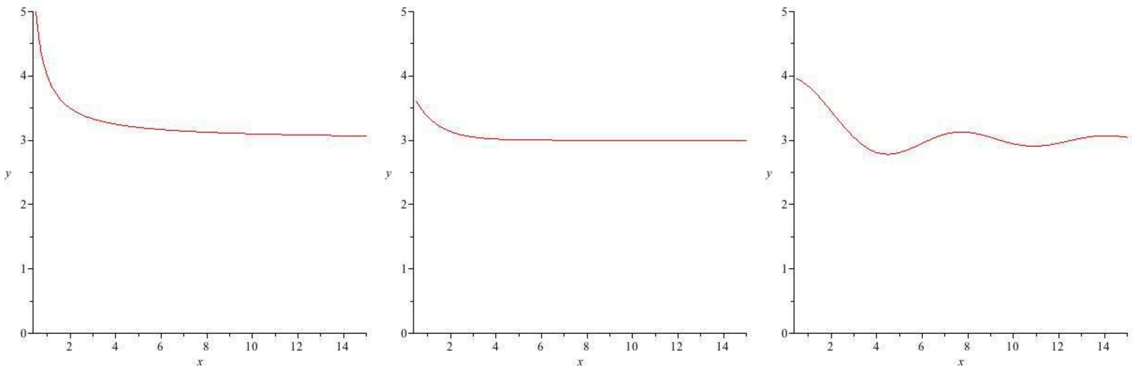
Intuitive
Pictorial
Formal
Computational

The book does a pretty good job on these, but most students do not learn limits all that well from a book so I am going to repeat some of that in these notes and in class. But please do read Section 2.2, 2.3, 2.4 and 2.6.

Intuitive: The limit as $x \to a$ of $f(x)$ is the value (if any) that $f(x)$ gets close to when x gets close to (but does not equal) a. This is denoted $\lim_{x\to a} f(x)$. If we only let x approach a from one side, say from the right, we get the one-sided limit $\lim_{x\to a^+} f(x)$.

Please observe the syntax: If I tell you a function f and a value a then the expression $\lim_{x\to a} f(x)$ takes on a value (perhaps "undefined" but nonetheless a value). The variable x is a bound or "dummy" variable; it does not have a value in the expression and does not appear in the answer; it stands for a continuum of possible values approaching a.

Pictorial: if the graph of f appears to zero in on a point (a, b) as the x-coordinate gets closer to a, then that is the limit (even if the actual point (a, b) is not on the graph). Look at Example 2 on page 67 to see what I mean about (a, b) not needing to be on the graph. We can take limits at infinity as well as at a finite number. The limit as $x \to \infty$ is particularly easy visually: if $f(x)$ gets close to a number C as $x \to \infty$ then f will have a horizontal asymptote at height C (if you allow the function to possibly cross the line and double back, and still call it an asymptote). Thus $3 + \frac{1}{x}$, $3e^{-x}$ and $3 + \frac{\sin x}{x}$ all have limit 3 as $x \to \infty$.



Formal: The precise definition of a limit is found in Section 2.3. An informal poll of last semester's students showed that zero out of 45 remembered covering this in their previous course (despite the fact it was on most syllabi). The good news is, we don't have to spend a lot of time on it. However, you should see it at least once, enough that you grasp it.

The formal definition makes the intuitive definition precise. The intuitive definition is that $\lim_{x \to a} f(x) = L$ if $f(x)$ gets close to L when x gets close to a. The formal definition formalizes "gets close to" and "when". The formal definition is

> $\lim_{x \to a} f(x) = L$ if for any small tolerance ε in the y value there is a corresponding small tolerance δ in the x value such that trapping the x value in the interval $[a - \delta, a + \delta]$ (but x is not allowed to equal a) forces the y value to be trapped in the interval $[L - \varepsilon, L + \varepsilon]$. In symbols:
>
> $$0 < |x - a| < \delta \Longrightarrow |f(x) - L| < \varepsilon\,.$$

Computational: There are five theorems in Section 2.2. You should know them all. I hope they seem intuitive to you but they may not. They are not too difficult however. Computation will be discussed mostly in the next lesson or in recitation (for example, the square root trick, which is useful for engineering but doesn't come up a whole lot in business applications).

Next, skip ahead to Section 4.5, page 256, and read L'Hôpital's rule. Actually you need to read the whole section: the next page tells you how to iterate L'Hôpital's rule and the page after that tells you how to deal with indeterminate forms other than 0/0. L'Hôpital's rule is useful and it is also pretty easy to use. The one thing you need to be careful of is trying to apply it when you don't have an indeterminate form to begin with. I put a bunch of questions on this in the pre-homework, which you can try once you've reviewed Section 4.5 of the text.

Finally, here is a trick that come in handy. When evaluating a limit with a power in it, try taking logs. If you can find the limit of the log, then exponentiate to get the limit of the original expression.

Example: To find $\lim_{x\to 0}(1+2x)^{1/(3x)}$ we first find the limit of the natural log: $\lim_{x\to 0}\ln[(1+2x)^{1/(3x)}] = \lim_{x\to 0}\ln(1+2x)/(3x) = \lim_{x\to 0}\frac{2/(1+2x)}{3}$ by L'Hôpital's rule. This evaluates to 2/3 at $x = 0$. The original limit is therefore $e^{2/3}$.

Examples of limit computations with rational functions (quotients of polynomials):

(i) Quotient of polynomials: compute $\lim\limits_{x\to\infty}\dfrac{x^2+3x}{2x^2-1} = \dfrac{1}{2}$.

You can read Example 20 on page 114 to see how to do this or wait till the next lesson when we use asymptotic equality to handle this type of limit at a glance.

(ii) Factoring out 0/0: compute $\lim\limits_{x\to 3}\dfrac{x-3}{x^2-9} = \dfrac{1}{6}$.

Everywhere except at $x = 3$ you can factor $x - 3$ out of the numerator and denominator. You obtain $\dfrac{1}{x+3}$ whenever $x \neq 3$, which has the limit (in fact the value) $\dfrac{1}{6}$ at $x = 3$.

Limits at infinity

You will notice in the worksheets and problems sets that most of our limits are taken at infinity. The formal definition is a little different but intuitively nothing changes. We say $\lim_{x\to\infty} f(x) = L$ if you can trap the y value within ε of L by trapping the x value above some number M. In symbols, for evey $\varepsilon > 0$ there has to be an M such that

$$x > M \Longrightarrow |f(x) - L| < \varepsilon \, .$$

Similarly, $\lim_{x\to-\infty} f(x) = L$ if for any ε you can trap y in $[L - \varepsilon, L + \varepsilon]$ by trapping x below some value $-M$:

$$x < -M \Longrightarrow |f(x) - L| < \varepsilon \, .$$

If $\lim_{x\to\infty} f(x)$ exists and is equal to the real number c, then the graph of f will approach the horizontal line $y = c$. This occurs, for example, when there is a horizontal asymptote at height c.

Limits of infinity.

I hope you've noticed that the statement that $f(x)$ has a limit as $x \to a$ is really the statement that $\lim_{x\to a} f(x) = L$ for some real number L. The other possibility is that the limit does not exist, for which you are free to use the abbreviation DNE. This can happen because the value becomes infinite or because it has a jump, or because it is too wiggly and never settles down. Under certain conditions we say the limit is $+\infty$. NOTE: THIS IS LINGUISTAICALLY VERY MISLEADING. It is a special case of DNE. Thus we can say simultaneoulsy that the limit is $+\infty$ and that the limit DNE.

To be precise, we say that $\lim_{x\to a} f(x) = +\infty$ if for any M we can trap the y value above M by trapping the x value near enough (but not equal to) a. Formally, $\lim_{x\to a} f(x) = \infty$ if for every M there is a $\delta > 0$ such that

$$0 < |x - a| < \delta \Rightarrow f(x) > M \, .$$

You can have a limit at $\pm\infty$ of $\pm\infty$. For example $\lim_{x\to\infty} f(x) = -\infty$ when for every M there is a constant B (I used B because we already used M) such that whenever $x > B$ we have $f(x) < M$.

One-sided limits.

I hope you read what's in the book about one-sided limits. We say $\lim_{x \to a^+} f(x) = L$ if you can trap $f(x)$ in any chosen interval $[L - \varepsilon, L + \varepsilon]$ by trapping x in an interval $(a, a+\delta)$ for suitably chosen δ. The intuition is that $f(x)$ gets near L when x gets near a approching from the right (on the number line) which is also called approaching from above (from higher numbers). Note again that you don't need to look at the value of f at a itself or at any point to the left of a. In fact the function doesn't have to be defined at these places, just on an interval (a, b) for some $b > a$.

Last remarks on limits:

Continuity: A function f is continuous at a point a if it has a (honest, two-sided) limit at a and this limit is equal to the function value at a (in other words, now we do require the point $(a, f(a))$ to be on the graph along with the other nearby values of f. Continuity is important later because it comes up in the hypotheses of theorems. You should read Section 2.5 and decide whether there is anything in there that is not intuitively obvious.

Limit of a sequence: The limit of a sequence $a_1, a_2, a_3, \ldots$ is nearly the same definition as $\lim_{x \to \infty} f(x)$ except that instead of a function f defined for every real value we have a term a_n that is defined only for integer values of n. Nevertheless, we say $\lim_{n \to \infty} a_n = L$ if and only if the terms of the sequence become arbitrarily close to L as n gets bigger.

2 Magnitudes and exponential behavior

2.1 Orders of growth

The topic of relative magnitudes is not covered in the book, though it has some overlap with Section 7.4. You can and should read Section 7.4 of the text, but I don't think the treatment there is adequate. Furthermore, they introduce some notions not commonly used, which we will avoid. Our goal is to make some vague statements more precise.

Why are we spending our time making a science out of vague statements? Answer: (1) people really think this way and it clarifies your thinking to make these thoughts precise; (2) a lot of theorems can be stated with these as hypotheses; (3) knowing the science of orders of growth helps to fulfill the Number Sense mandate because you can easily fit an unfamiliar function into the right place in the hierarchy of more familiar functions.

We focus on two notions in particular: when one function is **much** bigger/smaller/closer than another, and when two functions are **asymptotically equal**.

Comparisons at infinity

Mostly we will be comparing functions of x as $x \to \infty$. Let f and g be positive functions.

(*i*) We say the function f is **asymptotic to** the function g, short for "asymptotically equal to", if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1 \,.$$

(*ii*) The function f is said to be **much** smaller than g, or to grow "much more slowly" if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0 \,.$$

This is denoted $f \ll g$.

Example: True or false: $x \ll e^x$? It's true, in fact $x^n \ll e^x$ for any power n. In fact, $x^n \ll e^{x/M}$ for any n and M. Verbally,

> Exponential growth is faster than any power growth.

You might ask how we can be sure of this. For $x \ll e^x$ you can look at the graph, reason inductively, etc. In general, the easiest way to see it is to use L"ôpital's rule. Example 4(c) on page 258 of the text is easy to generalize. We'll discuss this a little when we talk about computation.

Example: True or false: $e^x + x \sim x$? Yes, as you can easily verify:

$$\lim_{x\to\infty} \frac{e^x + x}{e^x} = \lim_{x\to\infty} 1 + \frac{x}{e^x} = 1$$

because we have already seen that $\lim_{x\to\infty} x/e^x = 0$. Intuitively, $e^x + x$ is nearly identical to e^x because x is so small compared to e^x that it doesn't make a difference when you add it.

Discussion

This is a general rule: the function $g(x)+h(x)$ will be asymptotic to $g(x)$ exactly when $h(x) \ll g(x)$. Why? Because $(g(x) + h(x))/g(x)$ and $h(x)/g(x)$ differ by precisely 1, so their limits do, meaning that the limit of $h(x)/g(x)$ is zero if and only if the limit of $(g(x) + h(x))/g(x)$ is one.

This is very important when estimating. It allows you to clear away irrelevant terms: in any sum, every term that is much less than one of the others can be eliminated and the result will be asymptotic to what it was before.

Example: Find a nice function asymptotically equal to $\sqrt{x^2+1}$. The notion of "nice" is subjective, but I mean one you're comfortable with, can easily estimate, etc.

> Because $1 \ll x^2$ we can ignore the 1 and get $\sqrt{x^2}$ which is equal to x for all positive x. Therefore, $\sqrt{1+x^2} \sim x$.

It should be obvious that the relation $\sim$ is symmetric: $f \sim g$ if and only if $g \sim f$. Formally,

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1 \Longleftrightarrow \lim_{x\to\infty} \frac{g(x)}{f(x)} = 1$$

because one is the reciprocal of the other. On the other hand, the relation $f \ll g$ is not at all symmetric; in fact, it is not possible that simultaneously $f \ll g$ and $g \ll f$.

It is good to have an understanding of the relative sizes of common functions. Here are some basic facts, which we will come back to in the worksheets.

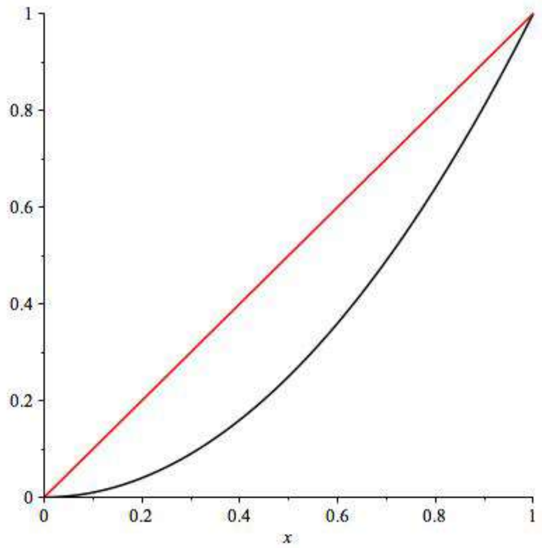
1. Positive powers all go to infinity but at different rates, with the higher power growing faster.
2. Exponentials grow at different rates and every exponential grows faster than every power.
3. Logarithms grow so slowly that any power of $\ln x$ is less than any positive power of x.

Comparisons elsewhere

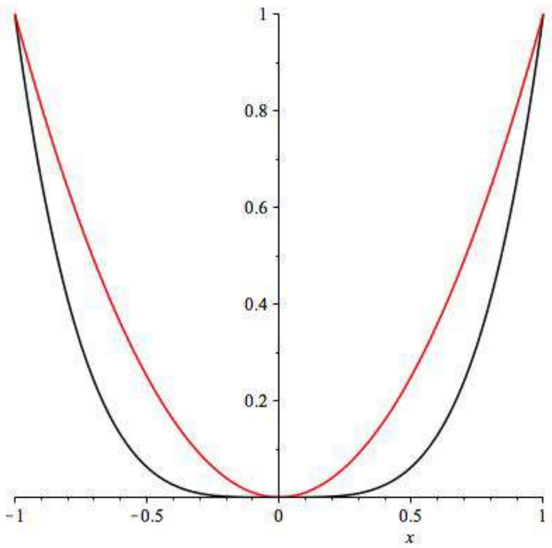
These same notions may be applied elsewhere simply by taking a limit as $x \to a$ instead of as $x \to \infty$. Usually this is done in order to compare how fast f and g go to either zero or infinity as $x \to a$. At a itself, the ratio of f to g might be $0/0$ or ∞/∞, which of course is meaningless, and can be made precise only by taking a limit as x approaches a. The notation, unfortunately, is not built to reflect whether $a = \infty$ or some other number. So we will have to spell out or understand by context whether the limit is intended to occur at infinity or some other location.

For example what about x and x^2? At infinity, we know $x \ll x^2$. But what about at zero, where both go to zero but a possibly different rates. Have a look at the graph on the next page. You can see that x has a postive slope whereas x^2 has a horizontal tangent at zero. Therefore, $x^2 \ll x$ as $x \to 0^+$. You can see it from the picture or from L'Hôpital:

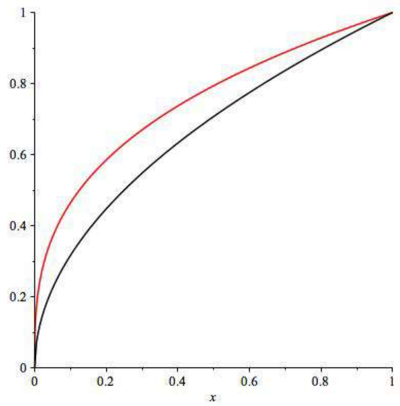
$$\lim_{x\to 0^+} \frac{x^2}{x} = \lim_{x\to 0^+} \frac{2x}{1} = 0\,.$$



Example: What about x^2 and x^4 near zero? Both have slope zero. By eye, x^4 is a lot flatter. Maybe $x^4 \ll x^2$ near zero. I don't think the picture settles it for sure, but L'Hôpital does.



Example: Compare $\sqrt{x}$ and $\sqrt[3]{x}$ near zero. Is one of these functions much smaller than the other as $x \to 0^+$? Here, the picture is pretty far from giving a definitive answer!



We try evaluating the ratio: $f(x)/g(x) = x^{1/2}/x^{1/3} = x^{1/2-1/3} = x^{1/6}$. Therefore,

$$\lim_{x\to 0^+} \frac{f(x)}{g(x)} = \lim_{x\to 0^+} x^{1/6} = 0$$

and indeed $x^{1/2} \ll x^{1/3}$. Intuitively, the square root of x and the cube root of x both go to zero as x goes to zero, but the cube root goes to zero a lot slower (that is, it remains bigger for longer).

"For sufficiently large x"

Often when discussing comparisons at infinity we use the term "for sufficiently large x". That means that something is true for every value of x greater than some number M (you don't necessarily know what M is). For example, is it true that $f \ll g$ implies $f < g$? No, but it implies $f(x) < g(x)$ for sufficiently large x. Any limit at infinity depends only on what happens for sufficiently large x.

Example: We have seen that $\ln x \ll \sqrt{x-5}$. It is not true that $\ln 6 < \sqrt{6-5}$ (the corresponding values are about 1.8 and 1) and it is certainly not true that $\ln 1 < \sqrt{1-5}$ because the latter is not even defined. But we can be certain that $\ln < \sqrt{x}$ for sufficiently large x. The crossover point is a little over 10.

2.2 Exponents and logarithms

While the section on exponents and logarithms is "math you already know," there is probably a fair amount of new learning for most of you. However, we will start with some algebraic identities that are purely review: you are expected already to know them and will be handling them by MML only. Please begin by reviewing these identities, which you will find in Sections 1.5 and 1.6 of the textbook. Page 37 in Section 1.5 has identities involving exponents, page 44–45 in Section 1.6 has identities involving logarithms, and your MML homework has a bunch of problems on these.

The notation we will use for logarithms in this class is: $\ln x$ for the natural log (base e); $\lg x$ for the base two logarithm, and $\log_b$ for a log to any other base. I may slip up and use log for ln because this is common in my research area. If I'm on the ball, you shouldn't see log without a base in this class.

The one most basic fact of all about logarithms is that $\log_b$ inverts the function $f(x) = b^x$. Therefore $\log_{10} 10^x = x$, meaning that you can tell something about $\log_{10} x$ just by knowing how many digits x has. If x has n digits then $\log_{10} x$ is between $n - 1$ and n.

Computation

Knowing just a few approximate values concerning logarithms allows you to do most computations without a calculator. On the next page, therefore, please find your very own Logarithm Cheat Sheet. Surprisingly, many important logs are good to within 1% even when you have only the first nontrivial decimal digit.

Logarithm Cheat Sheet

These values are accurate to within 1%:

$$
\begin{aligned}
e &\approx 2.7 \\
\ln(2) &\approx 0.7 \\
\ln(10) &\approx 2.3 \\
\log_{10}(2) &\approx 0.3 \\
\log_{10}(3) &\approx 0.48
\end{aligned}
$$

Some other useful quantities to with 1%:

$$
\begin{aligned}
\pi &\approx \frac{22}{7} \\
\sqrt{10} &\approx \pi \\
\sqrt{2} &\approx 1.4 \\
\sqrt{1/2} &\approx 0.7
\end{aligned}
$$

(ok so technically $\sqrt{2}$ is about 1.015% greater than 1.4 and 0.7 is about 1.015% less than $\sqrt{1/2}$)

Also useful sometimes: $\sqrt{3} = 1.732\ldots$ and $\sqrt{5} = 2.236\ldots$ both to within about 0.003%.

EXAMPLE: What is the probability of getting all sixes when rolling 10 six-sided dice? It's 1 in 6^{10} but how big is that? If we use base-10 logs, we see that $\log_{10}(6^{10}) = 10\log_{10} 6 = 10(\log_{10}(2) + \log_{10}(3)) \approx 10(.78) = 7.8$. So the number we're looking for is approximatelhy $10^{7.8}$ which is $10^7 \times 10^{0.8}$ or $10,000,000$ times a shade over $10^{.78}$ which is 6. So we're looking at a little over sixty million to one odds against.

This is not just a random example, it is always the best way to get a quick idea of the size of a large power. When the base is 10 we already know how many digits is has, but when the base is something else, we quickly compute $\log_{10}(b^a) = a \cdot \log_{10}(b)$.

EXAMPLE: Why is the 2.3 on your log cheatsheet so important? It converts back and forth between natural and base-10 logs. Remember, $\log_{10} x = \ln x / \ln 10$. Thus the constant $\ln 10$ is an inmportant conversion constant that just happens to be closer than it looks (the actual value is $2.302\ldots$). So for example,

$$e^8 \approx 10^{8/2.3} \approx 10^{3.5} = 100 \times 10^{0.5} \approx 3,000\,.$$

2.3 Exponential and logarithmic relationships

Much of what you learn on this topic will be a kind of insight into the nature of exponential and logarithmic functions:

- What happens to x when e^x doubles?
- Subtracting 5 from x does what to 2^x?
- If $y = Ax^b$ then how are $\ln x$ and $\ln y$ related?
- If we change the units of measurement, how does this affect the logarithm of the measurement?
- Why would we be more likely to compute the difference of the logarithms of measurements of two quantities than the logarithm of a single measure?

I will show you in a minute how to set up equations to answer this kind of question. But I should point out that some portion of what is to be learned is intuition. For example, take doubling: you know what doubling feels like, you have an intuitive feel that doubling $x + y$ is just like doubling x and y separately and adding them, and this is nearly independent of your knowledge of the distributive law. The knowledge I am trying to covey about exponential and logarithmic relationships will be imparted by two in-class activities (SEP 8 and 10) and a hands-on activity (SEP 11).

When answering a question about functional relationships, set up notation that distinguishes between before and after. For example in the first question above, use x_{before} and x_{after} to denote the two values of x, before and after. If that's too cumbersome then try something like x_0 and x_1. Once you choose the right notation, the problem almost solves itself. The information "e^x doubles" becomes the equation

$$e^{x_1} = 2e^{x_0} .$$

We are trying to capture the relation between x_0 and x_1 and it looks as if taking (natural) logs of both sides will get us there or nearly:

$$\ln\left(e^{x_1}\right) = \ln\left(2e^{x_0}\right)$$

simplifies to

$$x_1 = \ln 2 + x_0 .$$

So there's your answer: it corresponds to an increase by the (additive) amount of $\ln 2$ or roughly 0.7.

3 Sums and Integrals

Definite integrals are limits of sums. We will therefore begin our study of integrals by reviewing finite sums and the relation between sums and integrals. This will allow you to understand approximate values of integrals even when you can't evaluate the integral analytically (another instance of gaining number sense!). The first topic, finite sums, is very elementary but I don't know any good references so I'm including a reasonably complete treatment.

3.1 Finite sums

The preparatory homework for this sections deals with the nuts and bolts of writing finite sums. If given a sum such as $\sum_{n=5}^{19} \frac{3}{n-2}$ you should easily be able to tell what explicit sum it represents: how many terms, what are the first few and the last, how would you write it using an equation with ... and so forth. The above sum, for example, contains 15 terms and could be written as $\frac{3}{3} + \frac{3}{4} + \cdots + \frac{3}{17}$.

It is a little harder going the other way, writing a sum in Sigma notation when you are given its terms. One reason is that there is more than one way to do this. For example there is no reason why the index in the previous sum should go from 5 to 19. There have to be fifteen terms but why not write it with the index going from 1 to 15? Then it would look like

$$\sum_{n=1}^{15} \frac{3}{n+2} .$$

Another natural choice is to let the index run from 0 to 14:

$$\sum_{n=0}^{14} \frac{3}{n+3} .$$

All three of these formulas represent the exact same sum.

Another difficulty is that you need to know tricks to represent certain patterns with formulas. Really this is not a difficulty with smmations as much as with writing a formula to represent the general term a_n of a given sequence. Realize that these problems are inherently the same: writing the n^{th} term of a sequence as a function of n and writing the summand in a summation as a function of its index. The preparatory homework starts off with sequence writing and then has you do some summations as well.

Here are some tricks to write certain patterns. The term $(-1)^n$ bounces back and forth between $+1$ and -1, starting with -1 when $n = 1$ (or starting with $+1$ if your sum has a term for $n = 0$). You can incorporate this in a sum as a multiplicative factor and it will change the sign of every second term. Thus for example, to write the sum $1 - 2 + 3 - 4 + \cdots - 100$ you can write

$$\sum_{n=1}^{100} (-1)^{n+1} \cdot n \, .$$

Note that we used $(-1)^{n+1}$ rather than $(-1)^n$ so as to start off with a positive rather than a negative term.

When the sum has a pattern that takes a couple of steps to repeat, the greatest integer function can be useful. For example, $1+1+1+2+2+2+3+3+3+\cdots+10+10+10$ can be written as $\sum_{n=1}^{30} \left\lfloor \frac{n+2}{3} \right\rfloor$.

Sequences and sums can use definitions by cases just the way functions do. Suppose you want to define a sequence with an opposite sign on every third term, such as $-1, -1, 1, -1 - 1, 1, \ldots$. You can do this by cases as follows.

$$a_n = \begin{cases} -1 & n \text{ is not a multiple of } 3 \\ 1 & n \text{ is a multiple of } 3 \end{cases}$$

Although you will not be required to know this, you can use sophisticated tricks to avoid this kind of definition by cases. One way[1] is to use the greatest integer function:

$$a_n = (-1)^{\lfloor 2(n-1)/3 \rfloor} \, .$$

Notational observations: A sequence denoted $a_1, a_2, a_3, \ldots$ could just as easily be written as a function $a(1), a(2), a(3), \ldots$. The value of a term a_n is a function of the index n and there is no difference whether we write n as a subscript or as an argument.

[1]Another way is to use complex numbers, but you'll have to ask me about that separately if you're curious.

Series you can explicitly sum

We will learn to sum three kinds of series: arithmetic (accent on the third syllable) series, geometric series and telescoping series.

Arithmetic series

An arithmetic series is a sum in which the terms increase or decrease by the same amount (additively) each time. You can always write these in the form $a_n = A + dn$ where A is the initial term and d is how much each term increases over the one before (it could be negative if the terms decrease). Here you should start the sum at $n = 0$ or else use the term $A + (d - 1)n$. The standard trick for summing these is to pair up the first and last, the second and second-to-last, and so on, recognizing that each pair sums to twice the average and therefore that the sum is the number of terms times the average term. Here is an example in a particular case and then the general formula.

Example: Evaluate $\sum_{n=13}^{29} n$. There are 17 terms and the average is 21, which can be computed by averaging the first and last terms: $(13 + 29)/2 = 21$. Therefore, the sum is equal to $17 \times 21 = 357$.

General case: Evaluate $\sum_{n=0}^{M} A + dn$. There are $M+1$ terms and the average is $A + (dM/2)$. Therefore the sum is equal to $(M + 1)(A + (dM/2)) = A(M + 1) + dM(M + 1)/2$.

Geometric series

A geometric series is a sum in which the terms increase or decrease by the same multiplicative factor each time. You can always write these in the form $a_n = A \cdot r^n$ where A is the initial term and r is the factor by which the term increases each time. If the terms decrease then r will be less than 1. If they alternate in sign, r will be negative. Also, again, A will be the initial term only if one starts with the $n = 0$ term or changes the summand to $A \cdot r^{n-1}$.

The standard trick for summing these is to notice that the sum and r times the sum are very similar. I'll explain with an example.

EXAMPLE: Evaluate $\displaystyle\sum_{n=1}^{10} 7 \cdot 4^{n-1}$.

To do this we let S denote the value of the sum. We then evaluate $S - 4S$ (because $r = 4$). I have written this out so you can see the cancellation better.

$$\begin{aligned} S - 4S &= 7 + 28 + 112 + \cdots + 7 \cdot 4^9 \\ &\quad - (28 + 112 + \cdots + 7 \cdot 4^9 + 7 \cdot 4^{10}) \\ \\ &= 7 - 7 \cdot 4^{10}\,. \end{aligned}$$

From this we easily get $S = (7 - 7 \cdot 4^{10})/(1 - 4) = 7(4^{10} - 1)/3$.

GENERAL CASE: Evaluate $\displaystyle\sum_{n=1}^{M} A \cdot r^{n-1}$.

Letting S denote the sum we have $S - rS = A - Ar^n$ and therefore

$$S = A\,\frac{1 - r^n}{1 - r}\,.$$

Infinite series

No discussion of finite series would be complete without a mention of infinite series. There is a whole theory of convergence of infinite series that they teach in Math 104. Here we'll stick to what's practical. It should be obvious that $1 + 2 + 4 + \cdots$ does NOT converge, while $1/2 + 1/4 + 1/8 + \cdots$ DOES converge, and in fact converges to 1. There are eleven theorems and tests in the book about when series converge. From a practical point of view, all you need is two things: the definition, and an example.

> *Definition:* An infinite sum $\sum_{n=1}^{\infty} a_n$ is said to converge if and only if the partial sums $S_M = \sum_{n=1}^{M} a_n$ form a convergent sequence. In other words, if $\lim_{M\to\infty} S_M$ exists and is equal to L, then $\sum_{n=1}^{\infty} a_n$ is said to equal L.

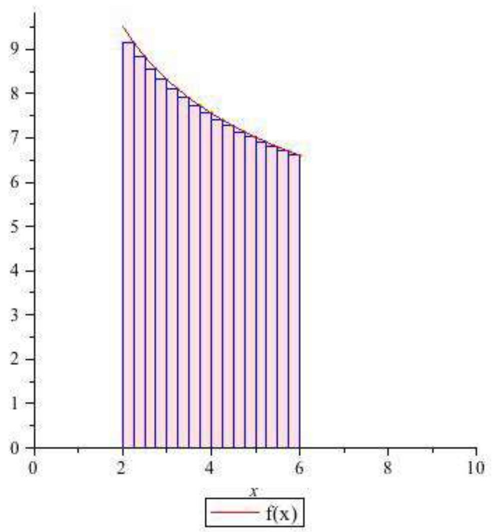
EXAMPLE: If $a_n = (1/2)^n$ then $S_M = 1 - (1/2)^M$. Clearly $\lim_{M\to\infty} S_M = 1$ so we say that $\sum_{n=1}^{\infty}(1/2)^n = 1$.

3.2 Riemann sums

In this unit we recap how areas lead to integrals and then, by the Fundamental Theorem of Calculus, to anti-derivatives.

Areas under graphs

Thankfully, Sections 5.1–5.3 do a nice job in explaining areas of regions under graphs as limits of areas of regions composed of rectangles. I will just point out the highlights. This figure shows a classical rectangular approximation to the region under a graph $y = f(x)$ between the x values of 2 and 6. The rectangular approximation is composed of 16 rectangles of equal width, all of which have their base on the x-axis and their top edge intersecting the graph $y = f(x)$. The rectangular approximation is clearly very near to the actual region, therefore the area of the region will be well approximated by the area of the rectangular approximation. This is easy to compute: just sum the width times height. The sum that gives this area is known as a Riemann sum.



Because the height is not constant over the little interval, there is no one correct height. You could certainly cover the targeted area with your rectangles by always choosing the highest point in each interval. That is called the **upper Riemann sum** (see page 300). If you go only as high as the least value of f in the interval, that is the **lower Riemann sum**, and these rectangles together will surely lie inside your targeted area. If one always chooses the top-left corner of the rectangle to lie on the graph then this is called the **left-Riemann sum**; if one always chooses the top-right corner of the rectangle to lie on the graph, this is called **right-Riemann sum**.

If f is increasing over the whole interval $[a, b]$ then a left-Riemann sum will also be a lower Riemann sum and a right-Riemann sum will be an upper Riemann sum; if f is decreasing, this correspondence is reversed. The example in the figure is of a right-Riemann sum, which is also a lower Riemann sum, with $a = 2$, $b = 4$, and a partition of the x-axis into 16 equal strips.

The definite integral is defined as such a limit. Specifically,

$$\int_a^b f(x)\,dx$$

is defined as the limit of the Riemann sums as the width of the rectangles goes to zero. So far we have not invoked the Fundamental Theorem of Calculus, so we are not connecting this with any kind of anti-derivative. We just have a definition of $\int_a^b f(x)\,dx$.

Interpretations other than area

Most people who compute integrals are not particularly interested in areas of regions. Integrals are interesting because the same math that computes the area of a region computes many other things as well. In general, it represents a total. If $f(t)$ is a quantity of something being delivered over time, such as water flow in gallons per minute, then $\int_a^b f(t)\,dt$ is the total amount delivered between time a and time b. If $f(t)$ is an acceleration then $\int_a^b f(t)\,dt$ is the total change in velocity from time a to time b.

The units of $\int_a^b f(x)\,dx$ are the units of f times the units of x. You can see this because the rectangles that make up the Riemann sum have units of height (units of f) times width (units of x). For example, suppose the x-axis is time (say hours) and the y-axis is number of people working at the given time; then the area is interpreted as person-hours of work (formerly known as man-hours). Thus $\displaystyle\int_a^b f(x)\,dx$ represents the total person-hours worked from time a to time b. If the y-axis represents a rate of change as the x-axis quantity changes, then the area represents total change. For example if x is time and y is velocity (rate of change of position with respect to time) then $\int_a^b f(x)\,dx$ is the total change in position from time a to time b. The units still come out right because velocity $\times$ time $=$ distance.

Fundamental Theorem of Calculus

The principles allowing us to evaluate integrals are these.

> (1) If the partition P is sufficiently fine, the upper Riemann sum U and the lower Riemann sum L will be very close. In fact the limit as P becomes finer of U and the limit as P becomes finer of L both exist and are equal to a number I which is, by definition, the value of the integral.
>
> (2) Amazingly, you can evaluate I exactly if you can find an anti-derivative F for the function f. The value of I will then be $F(b) - F(a)$.

The last part of this definition/theorem is a version of the Fundamental Theorem of Calculus. I suppose you already know it, but it's still very cool. Let's concentrate though on the other part. It says that we can use integrals to estimate sums or bound them and vice versa. In the next section we will discuss using U and L to get bounds. For now, we'll just say: if P is reasonably fine then any Riemann sum (U, L, or something in between) is pretty close to I.

A note on signed area: Area is defined to be a positive quantity. However, integrals compute *signed* area. Thus $\int_a^b f(x)\,dx$ computes the area between $x = a$ and $x = b$ below the graph of f but above the x-axis, with area below the x-axis counting as negative. Be careful about this, especially if there are both positive and negative pieces of the area.

Anti-derivatives

The FTC says areas are computed by anti-derivatives. Students from previous terms identified confusion as to exactly what an anti-derivative is. The **indefinite integral** has the notation $\int f(x)\,dx$ and represents any function whose derivative is the function $f(x)$. Let's say $F(x)$ is such a function.

The confusion lies when considering $F(x)$ as both a function of x and an integral. What integral has derivative equal to $f(x)$? Answer: $\int_a^x f(t)\,dt$. Note several things. (1) x appears as the upper limit of the integral. (2) we changed the name of the variable of integration to t. (3) the lower limit of integration can be any constant. It is (2) that is the most confusing: an integral from a constant to a variable is a function of that variable! If you're wondering why we changed x to t inside, it's to avoid confusion. The variable of integration is a **bound variable** also known as a **dummy variable**. It has no value, rather it is summed over. The value of the definite integral $\int_b^x f(t)\,dt$ depends on the values of b and x and the function f, but not on the value of t; there is no value of t. Please compare to pages 329–330 and Theorem 4 in the textbook.

3.3 Bounding and estimating integrals and sums

Both integrals and sums represent areas: an integral is the area under a curve and a sum is an area under a bunch of rectangles. You know one area is bigger than another when the first region completely covers the second region. Based on this, you can bound an integral by a sum or vice versa. To find a sum that is an upper bound for an integral, represent the integral as an area and find a sum whose area representation covers that of the integral. This is just the same as finding in upper Riemann sum. Similarly you can find a sum to give a lower bound for an integral, namely a lower Riemann sum. Going the other way, if you have a sum you can find an integral whose area completely covers that of the sum, which will give you an upper bound for the sum. Similarly, an integral whose area is completely contained in the rectangles for the sum will be a lower bound for the sum. We will practice this both ways: first, given an integral, bound it above and below by sums; secondly, given a sum, bound it above and below by integrals. At the very end of this section, we'll see how to get a good estimate for an integral that is neither an upper nor a lower bound (the trapezoidal estimate).

Estimating integrals using sums

The upper Riemann sum U is always an upper bound and the lower sum L is always a lower bound. When the function is monotone (either increasing or decreasing) then these are left- or right-Riemann sums and can therefore be computed routinely (though it may be tedious).

EXAMPLE: Find ten-term sums that are upper and lower bounds for $\int_1^2 \frac{1}{1+x^3}\,dx$.

The function $1/(1+x^3)$ is decreasing so the the left-Riemann sum (evaluate f at the left endpoint of each interval) is an always an upper sum and the right-Riemann sum is always a lower sum. These sums are easy to represent.

$$U = \sum_{j=0}^{9} f\left(1+\frac{j}{10}\right)\cdot\left(\frac{1}{10}\right) \tag{3.1}$$

$$L = \sum_{j=1}^{10} f\left(1+\frac{j}{10}\right)\cdot\left(\frac{1}{10}\right)$$

You can evaluate these by hand as 0.27430... and 0.2354... respectively [2].

Estimating sums using integrals

It is more interesting going the other way. Given a sum, how do we bound it by an integral? It's not hard to write an integral to which the sum is approximately equal, but to ensure that the integral lies above or below the sum we might have to do some fiddling. We use the fact that if a sum S is an upper Riemann sum for an integral I then I is a lower bound for S.

EXAMPLE: Find upper and lower bounds for the sum S_n defined by $\displaystyle\sum_{k=1}^{n} \frac{1}{k}$.

The lower bound is easy: if we put a rectangle of height $1/k$ above the interval $[k, k+1]$, for each k from 1 to n, then the union of rectangles is the upper Riemann sum for $\displaystyle\int_1^{n+1} \frac{1}{x}\, dx$. Thus $S \geq \ln(n+1)$ and we have found a lower bound for S. For an upper bound, one trick that works is to use fit all terms but the first of S underneath the graph of $1/x$ from 1 to n and then add the extra 1 from the first term. Thus $S \leq 1 + \int_1^n (1/x)\, dx = 1 + \ln(n)$. To summarize,

$$\ln(n+1) \leq S \leq 1 + \ln(n)\,.$$

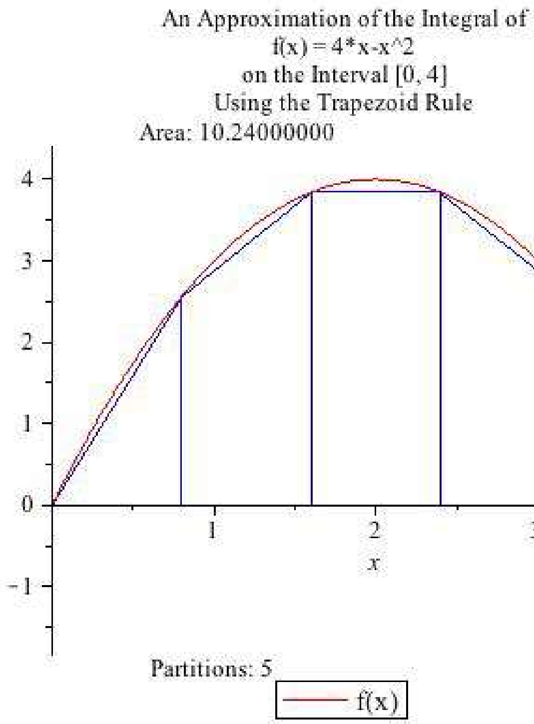
For $n = 50$ this comes out approximately as $3.93 \leq S \leq 4.92$.

Trapezoidal approximation

Sometimes it can be frustrating using Riemann sums because a lot of calculation doesn't get you all that good an approximation. You can see a lot of "white space" between the function f and the horzontal lines at the top of the rectangles that make up the upper or lower Riemann sum. If instead you let the rectangle become a right trapezoid, with both its top-left and top-right corner on the graph $y = f(x)$, then you get what is known as the **trapezoidal approximation**. The figure shows a trapezoidal approximation of an integral $\int_0^4 f(x)\, dx$ with five trapezoids. Note that the first and last trapezoid are degenerate, that is, one of the vertical sides has length

[2] If you are able to evaluate the integral exactly as $\pi/(6\sqrt{3}) + \ln(3/4)/6 \approx 0.25435...$ then you probably shouldn't be in this course.

zero and the trapezoid is actually a right triangle. It is perfectly legitimate for one or more of the trapezoids to be degenerate.



Because the tops of the slices are allowed to slant, they remain much closer to the graph $y = f(x)$ than do the Riemann sums. Because the area of a right trapezoid is the average of the areas of the two rectangles whose heights are the value of f at the two endpoints, it is easy to compute the trapezoidal approximation: it is just the average of the left-Riemann sum and the right-Riemann sum corresponding to the same partition into vertical strips.

Let's check what the trapezoidal approximation gives for the integral at the beginning of this section: $\int_1^2 \frac{1}{1+x^2}$. Adding the formlae for U and L and dividing by 2 yields

$$\frac{1}{2}\frac{f(1)}{10} + \frac{1}{2}\frac{f(0)}{10}\sum_1^9 \frac{1}{10} f\left(1 + \frac{j}{10}\right) .$$

In words, sum the values of f along a regular grid of x-values, counting endpoints as half, and multiply by the spacing between consecutive points.

The trapezoidal estimate is usually much closer than the upper or lower estimate, though it has the drawback of being neither an upper nor a lower bound. However, if you know the function to be concave upward then the trapezoidal estimate is an upper bound. Similarly if $f'' < 0$ on the interval then the trapezoidal estimate is an lower bound. In the figure, f is concave downward and the trapezoidal estimate is indeed a lower bound.

EXAMPLE: The function $1/(1+x^3)$ is concave upward on $[1, 2]$ (compute and see that the second derivative is a positive quantity divided by $(1 + x^3)^3$) so the trapezoidal estimate should be not only very close but an upper bound. Indeed, the trapezoidal estimate is the average of the upper and lower previously computed and is equal to 0.25485... which is indeed just slightly higher than the true value of 0.25425....

Part II

New topics in integration

4 Integration techniques

We are now out of Part I of the course, where everything goes back to number sense, and into a segment of the course that involves learning a skill. It's a high level skill, but you're good at that kind of thing or you wouldn't be here. So relax and enjoy some clean and satisfying computation. The material corresponds to Sections 5.5, 5.6 and 8.2 of the textbook. The first method ought to be review, and the second ought to be new, though due to your varied backgrounds some might find both to be review or both to be new.

Curricular note: in Math 104 they spend a whole lot of time on integration techniques, nearly half the course (rmember the slide I showed you of the Math 104 final?). These days you can get your computer (or even your Wolfram Alpha iPhone app) to do this for you so there isn't as great a need. But you need some familiarity in order to make sense of things, and learning the two most pervasive techniques strikes a reasonable balance.

4.1 Substitution

The most common way of doing a integral by substitution, and the only way for indefinite integrals, is as follows.

1. Change variables from x to u (hence the common name "u-substitution")
2. Keep track of the relation between dx and du
3. If you chose correctly you can now do the u-integral
4. When you're done, substitute back for x

The most common substitution is when you let $u = h(x)$ for some function h. Then $du = h'(x)\,dx$. Usually you don't do this kind of substitution unless there will be an $h'(x)\,dx$ term waiting which you can then turn into du. Also, you don't do this unless the rest of the occurences of x can also be turned into u. If h has an inverse function, you can do this by substituting $h^{-1}(u)$ for x everywhere. Now when you reach the fourth step, it's easier because you can just plug in $u = h(x)$ to get things back in terms of x.

Please read the examples in Section 5.5 – there are a ton. I will give just one.

EXAMPLE: Compute $\displaystyle\int \sin^n x \, \cos x \, dx$.

Solution: substitute $u = \sin x$ and $du = \cos x \, dx$. This turns the integral into $\int u^n \, du$ which is easily valuated as $u^{n+1}/(n+1)$. Now plug back in $u = \sin x$ and you get the answer

$$\frac{\sin^{n+1} x}{n+1} .$$

You might think to worry whether the substitution had the right domain and range, was one to one, etc., but you don't need to. When computing an indefinite integral you are computing an anti-derivative and the proof of correctness is whether the derivative is what you started with. You can easily check that the derivative of $\sin^{n+1} x/(n+1)$ is $\sin^n x \cos x$. There are a zillion examples of this in Section 5.5.

When evaluating a definite integral you can compute the indefinite integral as above and then evaluate. A second option is to change variables, including the limit of integration, and then never change back.

EXAMPLE: Compute $\displaystyle\int_1^2 \frac{x}{x^2+1} \, dx$.

If we let $u = x^2 + 1$ then $du = 2x \, dx$, so the integrand becomes $(1/2) \, du/u$. If x goes from 1 to 2 then u goes from 2 to 5, thus the integral becomes

$$\int_2^5 \frac{1}{2} \frac{du}{u} = \frac{1}{2} (\ln 5 - \ln 2) .$$

Of course you can get the same answer in the usual way: the indefinite integral is $(1/2) \ln u$; we substitute back and get $(1/2) \ln(x^2 + 1)$. Now we evaluate at 2 and 1 instead of 5 and 2, but the result is the same: $(1/2)(\ln 5 - \ln 2)$.

Some useful derivatives

A large part of exact integration is recognizing when something is a derivative of something familiar. Here is a list of functions whose derivatives you should stare at long enough to recognize if they come up. (Yes, you can put them on a cheatsheet when exam time comes.)

$$
\begin{aligned}
\frac{d}{dx}\tan x &= \sec^2 x \\
\frac{d}{dx}\sec x &= \sec x \tan x \\
\frac{d}{dx}\arcsin x &= \frac{1}{\sqrt{1-x^2}} \\
\frac{d}{dx}\arctan x &= \frac{1}{1+x^2} \\
\frac{d}{dx}\operatorname{arcsinh} x &= \frac{1}{\sqrt{1+x^2}} \\
\frac{d}{dx}\operatorname{arccosh} x &= \frac{1}{\sqrt{x^2-1}} \\
\frac{d}{dx}\operatorname{arctanh} x &= \frac{1}{1-x^2}
\end{aligned}
$$

4.2 Integration by parts

The integral by parts formula

$$\int u\,dv = uv - \int v\,du$$

is pretty well explained in Section 8.2 of the textbook. Here I will just mention a couple of the trickier instances of integration by parts.

Repeated integration by parts

As you will see, when one of the functions involved is e^x, and you take $dv = e^x\,dx$, then $v\,du$ will still have an e^x in it. In that case you can integrate by parts again. Will this ever stop? Well if the original u is a polynomial $p(x)$ then du will be $p'(x)\,dx$ so it will have degree one less, and if you repeat enough times you'll get to stop eventually.

Similarly, if $dv = \sin x\, dx$ or $\cos x\, dx$, then the v term will just cycle through sines and cosines and if it's multiplied by a polynomial, the degree will go down each time you integrate by parts and eventually you'll get an answer.

You could make an algorithm dealing will all integrals of the form $p(x)\, dx$. The book does exactly this and calls it **tabular integration**. If you want to learn to integrate $p(x)e^x$, $p(x)\sin x$ and $p(x)\cos x$ this way, go ahead. As far as I'm concerned, it is just as easy to do it out. Well if p has very high degree then probably you'll want to make up some kind of shortcut, but I don't insist that it be the book's version of tabular integration. I will give just one example because this is something the book handles well.

EXAMPLE: Integrate $\displaystyle\int (x^3 + 3x)\ \cos x\, dx$. Taking $u = x^3 + 3x$ and $dv = \cos x\, dx$ gives

$$\int (x^3 + 3x)\ \cos x\, dx = (x^3 + 3x)\sin x - \int (3x^2 + 3)\sin x\, dx\,.$$

Setting aside the first term on the right-hand side, we work on the second, integreating by parts again. There will be fewer double negatives if we take the minus sign inside and attach it to the $\sin x\, dx$ term:

$$\int (3x^2 + 3)(-\sin x)\, dx = (3x^2 + 3)\cos x - \int 6x\ \cos x\, dx\,.$$

One last integration by parts shows that the last integral is

$$6x \sin x - \int 6 \sin x dx = 6x \sin x + 6 \cos x\,.$$

So the whole thing comes out to be

$$(x^3 + 3x)\sin x + (3x^2 + 3)\cos x - (6x \sin x + 6\cos x)\,.$$

You can do all of this when there is a term like e^{3x} or $\cos(5x)$. Technically this is a substitution plus an integration by parts but when the substitution is just $5x$ for x, you can pretty much do it in your head. For example to integrate $\int xe^{5x}\, dx$ you can let $dv = e^{5x}\, dx$ and therefore $v = (1/5)e^{5x}$. The substitution is hidden in the correct evaluation of v from dv.

Back where you started but with a sign change

If you try to integrate $e^x \sin x$, you'll find you have a choice. You can make $u = e^x$ and $dv = \sin x \, dx$ or $u = \sin x$ and $dv = e^x \, dx$. Either way, if you do it twice, you're get back to where you started but with the opposite sign. That's good because you have something like

$$\int e^x \sin x \, dx = \text{otherstuff} - \int e^x \sin x \, dx \, .$$

Now you can move the last term to the right over to the left so it becomes twice the integral you want, and see that the integral you want is half of the other stuff on the right. Example 4 on page 464 of the textbook is a very clear description of this.

Last trick: you can always try $dv = dx$

When trying to integrate $\int f(x) \, dx$ it doesn't look like there's a u and a v but if you know the derivative of f you can always let $u = f$ and $dv = dx$ and get

$$\int f(x) \, dx = x f(x) - \int x f'(x) \, dx \, .$$

Whether this helps depends on whether the factor of x combines nicely with the f'. The easiest example that works out nicely is $\int \ln x \, dx$ which is Example 2 on page 463 of the textbook.

5 Integrals to infinity

Philosophy[3]

The only way we can talk about infinity is through limits.

EXAMPLE: we try to make sense of ∞/∞ or $\infty \cdot 0$ but there is no one rule for what this should be. When it comes up as a limit, such as $\lim_{x\to\infty} x^2/e^x$ then at least it is well defined. To evaluate the limit we need to use L'Hôpital's rule or some other means.

EXAMPLE: we try to make sense of $\sum_{n=1}^{\infty} a_n$. There is no already assigned meaning for summing infinitely many things. We **defined** this as a limit, which in each case needs to be evaluated:

$$\lim_{n\to\infty} \sum_{k=1}^{n} a_k \,.$$

It is the same when one tries to integrate over the whole real line. We define this as integrating over a bigger and bigger piece and taking the limit. In fact the definition is even pickier than that. We only let one of the limits of integration go to zero at a time. We define $\int_0^\infty f(x)\,dx$ to be $\lim_{M\to\infty} \int_0^M f(x)\,dx$. In general, for any lower limit b, we can define $\int_b^\infty f(x)\,dx$ to be $\lim_{M\to\infty} \int_b^M f(x)\,dx$. But if we want both limits to be infinite then we define the two parts separately. The value of $\int_{-\infty}^{\infty} f(x)\,dx$ is **defined** to equal

$$\lim_{M\to\infty} \int_b^M f(x)\,dx + \lim_{M\to-\infty} \int_M^b f(x)\,dx \,.$$

If either of these limits does not exist then the whole integral is defined not to exist. At this point you should be bothered by three questions.

1. What is b? Does it matter? How do you pick it?
2. If we get $-\infty + \infty$, shouldn't that possibly be something other than "undefined"?
3. Why do we have to split it up in the first place?

[3] You can skip this section if you care about computing but not about meaning.

The answer to the first question is, pick b to be anything, you'll always get the same answer. That's because if I change b from, say, 3 to 4, then the first of the two integrals loses a piece: $\int_3^4 f(x)\,dx$. But the second integral gains this same piece, so the sum is unchanged. This is true even if one or both pieces is infinite. Adding or subtracting the finite quantity $\int_3^4 f(x)\,dx$ won't change that.

The answer to the second question is that this is really our choice. If we allow infinities to cancel, we have to come up with some very careful rules. If you would like to study this sort of thing, consider being a Math major and taking the Masters level sequence Math 508–509. For the rest of us, we'll avoid it. This will help to avoid the so-called re-arrangement paradoxes, where the same quantity sums to two different values depending on how you sum it.

The last question is also a matter of definition. Consider the sign function

$$f(x) = \text{sign}(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

On one hand, $\int_{-M}^{M} f(x)\,dx$ is always zero, because the postive and negative parts exactly cancel. On the other hand, $\int_b^\infty f(x)\,dx$ and $\int_{-\infty}^b f(x)\,dx$ are always undefined. Do we want the answer for the whole integral $\int_{-\infty}^\infty f(x)\,dx$ to be undefined or zero? There is no intrinsically correct choice here but it is a lot safer to have it undefined. If it has a value, one could make a case for values other than zero by centering the integral somewhere else, for example $\int_{7-M}^{7+M} f(x)\,dx$ is always equal to 14.

5.1 Type I Improper integrals and convergence

The central question of this section is: how do we tell whether a limit such as $\int_b^\infty f(x)\,dx$ exists, and if so, what the value is?

Case 1: you know how to compute the definite integral

Suppose $\int_b^M f(x)\,dx$ is something for which you know how to compute an explicit formula. The formula will have M in it. You have to evaluate the limit as $M \to \infty$. How do you do that? There is no one way, but that's why we studied limits before.

Apply what you know. What about b, do you have to take a limit in b as well? I hope you already knew the answer to that. In this definition, b is any fixed number. You don't take a limit.

Here are some cases you should remember.

Type of integral	Condition for convergence
$\int_b^\infty e^{kx}\,dx$	
$\int_b^\infty x^p\,dx$	
$\int_b^\infty \frac{(\ln x)^q}{x}\,dx$	

You will work out these cases in class: write each as a limit, evaluate the limit, state whether it converges, which will depend on the value of the parameter, k, p or q. Go ahead and pencili them in once you've done this. The second of these especially, is worth remembering because it is not obvious until you do the computation where the break should be between convergence and not.

Case 2: you don't know how to compute the integral

In this case you can't even get to the point of having a difficult limit to evaluate. So probably you can't evaluate the improper integral. But you can and should still try to answer whether the integral has a finite value versus being undefined. This is where the comparison tests come in. You buildup a library of cases where you do know the answer (Case 1) and then for the rest of functions, you try to compare them to functions in your library.

Sometimes a comparison is informative, sometimes it isn't. Suppose that f and g are positive functions and $f(x) \leq g(x)$. Consider several pieces of information you might have about these functions.

(a) $\int_b^\infty f(x)\,dx$ converges to a finite value L.

(b) $\int_b^\infty f(x)\,dx$ does not converge.

(c) $\int_b^\infty g(x)\,dx$ converges to a finite value L.

(d) $\int_b^\infty g(x)\,dx$ does not converge.

In which cases can you conclude something about the other function? We are doing this in class. Once you have the answer, either by working it out yourself or from the class discussion, please pencil it in here so you'll have it for later reference. This is essentially the **direct comparison test** at the bottom of page 510 of the textbook.

Even better comparison tests

Here are two key ideas that help your comparison tests work more of the time, based on the fact that the question "convergence or not?" is not sensitive to certain things.

(1) It doesn't matter if $f(x) \leq g(x)$ for every single x as long as the inequality is true from some point onward. For example, if $f(x) \leq g(x)$ once $x \geq 100$, then you can apply the comparison test to compare $\int_b^\infty f(x)\,dx$ to $\int_b^\infty g(x)\,dx$ as long as $b \geq 100$. But even if not, once you compare $\int_{100}^\infty f(x)\,dx$ to $\int_{100}^\infty g(x)\,dx$, then adding the finite quantity $\int_b^{100} f(x)\,dx$ or $\int_b^{100} g(s)\,ds$ will not change whether either of these converges.

(2) Multiplying by a constant does not change whether an integral converges. That's because if $\lim_{M\to\infty} \int_b^M f(x)\,dx$ converges to the finite constant L then $\lim_{M\to\infty} \int_b^M Kf(x)\,dx$ converges to the finite constant KL.

Putting these two ideas together leads to the conclusion that if $f(x) \leq Kg(x)$ from some point onward and $\int_b^\infty g(x)\,dx$ converges, then so does $\int_b^\infty f(x)\,dx$. The theorem we just proved is:

> *If f and g are positive functions on some interval (b, ∞) and if there are some constants M and K such that*
>
> $$f(x) \leq Kg(x) \text{ for all } x \geq K$$
>
> *then convergence of the integral $\int_b^\infty g(x)\,dx$ implies convergence of the integral $\int_b^\infty f(x)\,dx$.*

We used to teach this as the main theorem in this section but students said it was too hard because of the phrase "there exist constants k and M." What are K and M, they would ask, and how do we find them? You can ask about this if you want, but don't worry, it's not required. Instead, just remember, if f is less than any multiple of g from some point on, you can use the comparison test, same as if $f \leq g$.

EXAMPLE: $f(x) = \dfrac{1}{3e^x - 5}$. Actually $f(x) \sim 3e^{-x}$ but all we need to know is that $f(x) \leq 4e^{-x}$ once x is large enough (in this case large enough so that $3e^x \geq 20$). We know that $\int^\infty e^{-x}\,dx$ converges, hence $\int^\infty f(x)\,dx$ converges as well.

5.2 Probability densities

This section is well covered in the book. It is also long: eleven pages. However, it is not overly dense. I will expect you to get most of what you need out of the textbook and just summarize the highlights. I will cover the philosophy in lecture (questions like, "What is probability really?") and stick to the mathematical points here.

A nonnegative continuous function f on a (possibly infinite) interval is a **probability density function** if its integral is 1. If we make a probability model in which some quantity X behaves randomly with this probability density, it means we believe the probability of finding X in any smaller interval $[a, b]$ will equal $\int_a^b f(x)\,dx$. Often the model tells us the form of the function f but not the multiplicative constant. If we know that $f(x)$ should be of the form Cx^{-3} on $[1, \infty)$ then we would need to find the right constant C to make this a probability density, meaning that it makes $\int_1^\infty Cx^{-3}\,dx$ equal to 1.

Several quantities associated with a probability distribution are defined in the book: mean (page 520), variance (page 522), standard deviation (page 522) and median (page 521). Please know these definitions! I will talk in class about their interpretations (that's philosphy again).

There are a zillion different functions commonly used for probability densities. Three of the most common are named in the Chapter: the exponential (page 521), the uniform (page 523) and the normal (page 524). It is good to know how each of these behaves and in what circumstances each would arise as a model in an application.

The exponential distribution

The exponential distribution has a parameter μ which can be any positive real number. Its density is $(1/\mu)e^{-x/\mu}$ on the positive half-line $[0, \infty)$. This is obviously the same as the density Ce^{-Cx} (just take $C = 1/\mu$) but we use the parameter μ rather than C because a quick computation shows that the mean of the distribution is equal to μ: integrate by parts with $u = x$ and $dv = \mu^{-1}e^{-x/\mu}$ to get

$$\int_0^\infty \frac{x}{\mu} e^{-x/\mu}\, dx = -xe^{-x}\Big|_0^\infty + \int_0^\infty e^{-x/\mu}\, dx = 0 + \left(-\mu e^{-x/\mu}\right)\Big|_0^\infty = \mu\,.$$

Note that when we evaluate these quantities at the endpoints zero and infinity we are really taking a limit for the infinite endpoint.

The exponential distribution has a very important "memoryless" propoerty. If X has an exponential density with any parameter and is interpreted as a waiting time, then once you know it didn't happen by a certain time t, the amount of further time it will take to happen has the same distribution as X had originally. It doesn't get any more or any less likely to happen in the the interval $[t, t+1]$ than it was originally to happen in the interval $[0, 1]$.

The median of the exponential distribution with mean μ is also easy to compute. Solving $\int_0^M \mu^{-1}e^{-x/\mu}\, dx = 1/2$ gives $M = \mu \cdot \ln 2$. When X is a random waiting time, the interpretation is that it is equally likely to occur before $\ln 2$ times its mean as after. So the median is significantly less than the mean.

Any of you who have studied radioactive decay know that each atom acts randomly and independently of the others, decaying at a random time with an exponential distribution. The fraction remaining after time t is the same as the probability that each individual remains undecayed at time t, namely $e^{-t/\mu}$, so another interpretation for the median is the **half-life**: the time at which only half the original substance remains.

The uniform distribution

The uniform distribution on the interval $[a, b]$ is the probability density whose density is a constant on this interval: the constant will be $1/(b-a)$. This is often thought of the least informative distribution if you know that the the quantity must be between the values a and b. The mean and median are both $(a + b)/2$. Example 11 on page 523 of the book discusses why the angle of a spinner should be modeled by a uniform random variable.

The normal distribution

The normal density with mean μ and standard deviation σ is the density

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} .$$

The **standard normal** is the one with $\mu = 0$ and $\sigma = 1$. There is a very cool mathematical reason for this formula, which we will not go into. When a random variable is the result of summing a bunch of smaller random variables all acting independently, the result is usually well approximated by a normal. It is possible (though very tricky) to show that the definite integral of this density over the whole real line is in fact 1 (in other words, that we have the right constant to make it a probability density).

Annoyingly, there is no nice antiderivative, so no way in general of computing the probability of finding a normal between specified values a and b. Because the normal is so important in statistical applications, they made up a notation for the antiderivative in the case $\mu = 0, \sigma = 1$:

$$\Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx .$$

So now you can say that the probability of finding a standard normal between a and b is exactly $\Phi(b) - \Phi(a)$. In the old, pre-computer days, they published tables of values of Φ. It was reasonably efficient to do this because you can get the antiderivative F of any other normal from the one for the standard normal by a linear substition: $F(x) = \Phi((x - \mu)/\sigma)$. Please be sure to read Example 13 on page 525 where this is explained in considerable detail.

5.3 Type II improper integrals

A type II improper integral occurs if we try to integrate $\int_a^b f(x)\, dx$ but somewhere on the interval $[a, b]$ the function f becomes discontiuous. You may not have realized at the time but the definition of the definite integral required f to be continuous on the interval over which you integrate. The most common way that f might fail to be continuous is if it becomes unbounded (e.g., goes to infinity) in which case you can imagine that the Riemann sums defining the integral could be very unstable (for example, there is no upper Riemann sum if the function f does not have a finite maximum).

Here are some examples: 1) integrating $p(x)/q(x)$ on an interval where q has a zero; 2) integrating $\ln(x)$ on an interval containing zero; 3) integrating $\tan x$ on an interval containing $\pi/2$.

Again, the way we handle this is to integrate only over intervals where the function is continuous, then take limits to approach the bad value(s). If the bad value occurs at the endpoint of the interval of integration, it is obvious how to take a limit. Suppose, for example, that f is discontinuous at b. Then define

$$\int_a^b f(x)\,dx := \lim_{c \to b^-} \int_a^c f(x)\,dx\,.$$

Note that this is a one-sided limit. We are not interested in letting c be a little bigger than b, only a little smaller. Similarly, if the discontinuity is at the left endpoint, a, we define

$$\int_a^b f(x)\,dx := \lim_{c \to a^+} \int_c^b f(x)\,dx\,.$$

Notice in both cases I have used the notation ":=" for "is defined as", to emphasize that this is a definition.

If there is a single value c in the interior of the interval, at which f becomes discontinuous, then $\int_a^b f(x)\,dx$ is defined by breaking into two integrals, one from a to c and one from c to b. Each of these has a discontinuity at an endpoint, which we have already discussed how to handle, and we then add the two results. Again, if either one is undefined, then the whole thing is undefined.

$$\int_a^b f(x)\,dx := \lim_{s \to c^-} \int_a^s f(x)\,dx + \lim_{s \to c^+} \int_s^b f(x)\,dx\,.$$

If there is more than one bad point, then we have to break into more than two intervals.

We do the same thing for testing convergence of Type II improper integrals as we did for Type I, namely we find a bunch that we can evaluate exactly and for the rest we compare to one of these. Again, the most useful cases turn out to be powers x^p, with $p = -1$ being the borderline case. One then has to learn the art of comparing to powers. Specifically, if you know p for which $f(x) \sim |x-a|^p$ as $x \to a$, or even that $f(x) \sim C|x-a|^p$ as $x \to a$, then you will be able to determine convergence. We will do some work on this in class but you may also want to check out Examples 4, 5, and 7 on pages 508–511 of the textbook.

Part III

Differential equations and Taylor series

6 Taylor Polynomials

The textbook covers Taylor polynomials as a part of its treatment of infinite series (Chapter 10). We are spending only a short time on infinite series (the next unit, Unit 7) and will therefore learn Taylor polynomials with a more direct, hands-on approach. Accordingly, the readings in the coursepack will be more central, I will be providing a bit more in terms of lecture, the pre-homework will be relatively short, with extra length in the regular homework devoted to problems that would normally be in the pre-homework.

6.1 Taylor polynomials

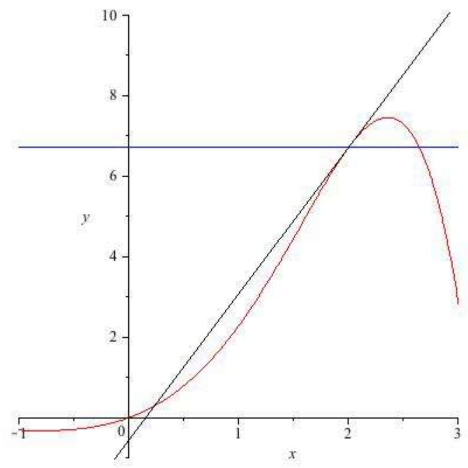
Idea of a Taylor polynomial

Polynomials are simpler than most other functions. This leads to the idea of approximating a complicated function by a polynomial. Taylor realized that this is possible provided there is an "easy" point at which you know how to compute the function and its derivatives. Given a function $f(x)$ and a value a, we will define for each degree n a polynomial $P_n(x)$ which is the "best n^{th} degree polynomial approximation to $f(x)$ near $x = a$."

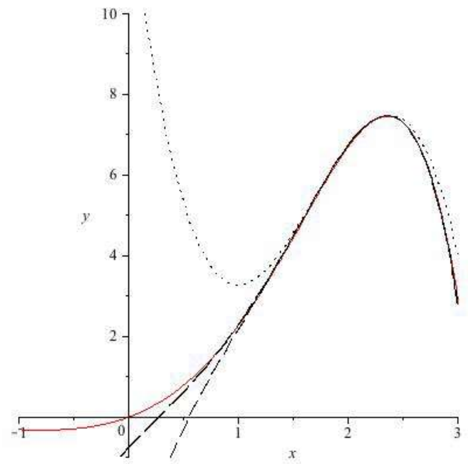
It pays to start very simply. A zero-degree polynomial is a constant. What is the best constant approximation to $f(x)$ near $x = a$? Clearly, the constant $f(a)$. What is the best linear approximation? We already know this, and have given it the notation $L(x)$. It is the tangent line to the graph of $f(x)$ at $x = a$ and its equation is $L(x) = f(a) + f'(a)(x - a)$. So now we know that

$$\begin{aligned} P_0(x) &= f(a) \\ P_1(x) &= f(a) + f'(a)(x - a) \end{aligned}$$

The figure on the next page shows the graph of a function f along with its zeroth and first degree Taylor polynomials at $x = 2$. The zeroth degree polynomial is the flat line and the first degree Taylor polynomial is the tangent line.



Just one more idea is needed to bust this wide open, that is to figure out $P_n(x)$ for all n: the polynomial $P_n(x)$ matches all the derivatives of f at a up to the n^{th} derivative. Check: P_0 matches the zeroth derivative, that is the function value, and P_1 matches the first derivative because both P_1 and f have the same first derivative at a, namely $f'(a)$. The next figure shows P_3, P_4 and P_5 at $x = 2$ for the same function, with P_5 shown in long dashes, P_4 in shorter dashes and P_3 in dots. As n grows, notice how P_n beceoms a better approximation and stays close to f for longer.



Taylor's formula

Using what we just said you can solve for what quadratic term is needed to match the second derivative. We used to make students go through this derivation but it took a lot of time and the students did not seem to feel it increased their understanding. Therefore, we will jump straight to the formula.

The definition uses some possibly unfamiliar notation: $f^{(k)}$ refers to the k^{th} derivative of the function f. This is better than f', f'', etc., because we can use it in a formula as k varies. $f^{(0}$ denotes f itself.

> **Definition of Taylor polynomial:** Let a be any real number and let f be a function that can be differentiated at least n times at the point a. The **Taylor polynomial** for f of order n about the point a is the polynomial $P_n(x)$ defined by
>
> $$P_n(x) := \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k \, .$$

Remember to read this sort of thing slowly. Here is roughly the thought process you should go through when seeing this for the first time.

- It looks as if P_n is a polynomial in the variable x with $n+1$ terms.
- When $a = 0$ it's a little simpler:

  $$P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k \, .$$

  The coefficients are the derivatives of f at zero divided by successive factorials.
- Hey, what's zero factorial? Oh, it's defined to be 1. Who knew?
- The degree of $P_n(x)$ will be n unless the coefficient on the highest power $(x-a)^n$ is zero, in which case the degree will be less.

Next you should try a simple example.

EXAMPLE: $f(x) = x$, $n = 3$ and $a = 2$. The value of $f(a)$ is 2 and the first three derivatives of $f(x)$ are constants: 1, 0, 0. Therefore

$$P_3(x) = 2 + 1 \cdot (x-2) + \frac{0}{2!}(x-2)^2 + \frac{0}{3!}(x-2)^3 \,.$$

In other words, $P_3(x) = x$. Obviously P_4, P_5 and so on will also be x. Maybe this example was too trivial. But it does point out a fact: if f is a polynomial of degree d then the terms of the Taylor polynomial beyond degree d vanish because the derivatives of f vanish. In fact, $P_n(x) = f(x)$ for all $n \geq d$.

EXAMPLE: $f(x) = e^x$, $n = 3$ and $a = 0$. We list the function and its derivatives out to the third one.

k	$f^{(k)}(x)$	$f^{(k)}(a)$	$\frac{f^{(k)}(a)}{k!}(x-a)^k$
0	e^x	1	1
1	e^x	1	x
2	e^x	1	$\frac{x^2}{2}$
3	e^x	1	$\frac{x^3}{6}$

Summing the last column we find that $P_3(x) = 1 + x + x^2/2 + x^3/6$.

EXAMPLE: Let $f(x) = \ln\sqrt{x}$ and expand around $a = 1$. We'll do the first two terms this time.

k	$f^{(k)}(x)$	$f^{(k)}(a)$	$\frac{f^{(k)}(a)}{k!}(x-a)^k$
0	$\ln\sqrt{x}$	0	0
1	$\frac{1}{2x}$	$\frac{1}{2}$	$\frac{1}{2}(x-1)$
2	$\frac{-1}{2x^2}$	$-\frac{1}{2}$	$-\frac{1}{4}(x-1)^2$

Summing the last column we find that $P_2(x) = \dfrac{x-1}{2} - \dfrac{(x-1)^2}{4}$.

6.2 Computing Taylor polynomials

You can always compute a Taylor polynomial using the formula. But sometimes the derivatives get messy and you can save time and mistakes by building up from pieces. Taylor polynomials follow the usual rules for addition, multiplication and composition. If f and g have Taylor polynonmials P and Q of order n then $f+g$ has Taylor polynomial $P+Q$. This is easy to see because the derivative is just the sum of the derivatives. Furthermore, the order n Taylor polynomial for fg is $P \cdot Q$ (ignore terms of order higher than n). This is because the product rule for the derivative of fg looks exactly like the rule for multiplying polynomials. I won't present a proof here but you can feel free to use this fact.

EXAMPLE: What is the cubic Taylor polynomial for $e^x \sin x$? The respective cubic Taylor polynomials are $1 + x + x^2/2 + x^3/6$ and $x - x^3/6$. Multiplying these and ignoring terms with a power beyond 3 we get

$$P_3(x) = x\left(1 + x + \frac{x^2}{2}\right) - \frac{x^3}{6} \cdot 1 = x + x^2 + \frac{x^3}{3}\,.$$

You can do the same thing with division, assuming you learned polynomial long division (this is useful? Who knew!).

Perhaps the most useful manipulation is composition. I will illustrate this by example. The Taylor polynomial for e^{x^2} is obtained by plugging in x^2 for x in the Taylor polynomial or series for $e^x : 1 + (x^2) + (x^2)^2/2! + \cdots$.

One last trick arises when computing the Taylor series for a function defined as an integral. Suppose $f(x) = \int_b^x g(t)\,dt$. Then $f'(x) = g(x)$ so if you know g and its derivatives, you know the derivatives of f. If g has no nice indefinite integral, then you don't know the value of f itself, except at one point, namely $f(b) = 0$. Therefore, a Taylor series at b is the most common choice for a function defined as $\int_b^x$ of another function.

EXAMPLE: Suppose $f(x) = \int_1^x \sqrt{1+t^3}\,dt$. The Taylor series can be computed about the point $a = 1$. From $f'(x) = \sqrt{1+x^3}$, $f''(x) = 3x^2/(2\sqrt{1+x^3})$ we get

$$f(1) = 0, \quad f'(1) = \sqrt{2}, \quad f''(1) = 3/(2\sqrt{2})$$

and therefore $P_2(x) = \sqrt{2}(x-1) + \dfrac{3}{4\sqrt{2}}(x-1)^2$.

Using Taylor polynomials to approximate

In the subsequent sections and in my lectures you will see where Taylor polynomials come from, why they are good approximations to the functions that generate them. You will also see precise statements about how close they are. For now though, we will take this on faith and see how to use them. In case this bothers you I will point out two quick things. (1) The Taylor polynomial of degree 0 is the constant $f(a)$. Surely this is a reasonable, if trivial, approximation to the function $f(x)$ when x is near a. (2) The Taylor polynomial of degree 1 is the linearization $f(a)+f'(a)\cdot(x-a)$. Again, you should already believe that this is a good approximation to $f(x)$ near $x=a$, in fact it is the best possible approximation by a linear function.

Example: What's a good approximation to $e^{0.06}$? A Taylor polynomial at $a=0$ will provide a very accurate estimate with only a few terms. The linear approximation 1.06 is already not bad. The quadratic approximation is

$$1+0.06+(1/2)(0.06)^2 = 1+0.06+0.0018 = 1.0618\,.$$

The true value is 1.0618365... so the quadratic approximation is quite good!

Taylor series are particularly useful in approximating integrals when you can't do the integral. Remember the problem of approximating $\int_0^{1/2} \cos(\pi x^2)\,dx$? It was not se easy to get a good answer with a trapezoidal appoximation. We can do better approximatng cos by a Taylor polynomial around $a=0$. You can directly compute that the first three derivatives are zero, or you can compute P_4 in one easy step like this: for the function $\cos x$, $P_2(x) = 1-x^2/2$; now plug in πx^2 for x to get $1-\pi^2x^4/2$. This is P_4. The nice thing about polynomials is that you can always integrate them. In this case,

$$\int_0^{1/2} P_4(x) = \left(x - \frac{\pi^2}{10}x^5 \right)\Bigg|_0^{1/2} .$$

This comes out to $1/2 - \pi^2/320 \approx 0.46916$ which is accurate to within 0.001.

6.3 Taylor's theorem with remainder

The central question for today is, how good an approximation to f is P_n? We will give a rough answer and then a more precise one.

Rough answer: $P_n(x) - f(x) \sim c(x-a)^{n+1}$ near $x = a$. For example, the linear approximation P_1 is off from the actual value by a quadratic quantity $c(x-a)^2$. If x differs from a by about 0.1 then $P_1(x)$ will differ from $f(x)$ by something like 0.01. If x agrees with a to four decimal places, then $P_1(x)$ should agree with $f(x)$ to about eight places. Similarly, the quadratic approximation P_2 differs from f by a multiple of $(x-a)^3$, and so on.

You can skip the justification of this answer, but I thought I'd include the derivation for those who want it because it's just an application of L'Hôpital's rule. Once you guess that $P_n(x) - f(x) \sim c(x-a)^n$, you can verify it by starting with the equation

$$\lim_{x \to 0} \frac{f(x) - P_n(x)}{(x-a)^{n+1}},$$

and repeatedly applying L'Hôpital's rule until the denominator is not zero at $x = a$. Because the derivatives of f and P_n at zero match through order n, it takes at least $n+1$ derivatives to get something nonzero, at which point the denominator has become the nonzero constant $(n+1)!$. The limit is therefore $f^{(n+1)}(a)/(n+1)!$, which may or may not be zero but is surely finite.

We know the Taylor polynomial is an order $(x-a)^{n+1}$ approximation but there is a constant c in the expression which could be huge. What about actual bounds can we obtain on $f(x) - P_n(x)$? These are given by Answer # 2, which is called Taylor's Theorem with Remainder.

> **Taylor's Theorem with Remainder:** Let f be a function with $n+1$ continuous derivatives on and interval $[a, x]$ or $[x, a]$ and let P_n be the order n Taylor polynomial for f about the point a. Then
>
> $$f(x) - P_n(x) = \frac{f^{(n+1)}(u)}{(n+1)!}(x-a)^{n+1}$$
>
> for some u between a and x.

The theorem is telling us that the constant c in the rough answer is equal to $f^{(n+1)}(u)/(n+1)!$ for this unknown u. This is at first a little mysterious and difficult to use, which is why we'll be doing some practice. The exact value of u will depend on a, x, n and f and will not be known. However, it will always be between a and x. This means we can often get bounds. We might know, for example, that $f^{(n+1)}$ is always positive on $[a, x]$ and is greatest at a, which would lead to

$$P_n(x) \leq x \leq P_n(x) + \frac{f^{(n+1)}(a)}{(n+1)!}(x-a)^{n+1}.$$

Example: Let $f(x) = e^{-x}$, $a = \ln 10$ and $n = 1$. How well does $P_2(x) = \frac{1}{10} - \frac{1}{10}(x - \ln 10)$ approximate e^{-x} for $x = \ln 10 + 0.2 \approx 2.502$? The remainder $R = e^x - P_n(x)$ will equal $f''(u)/2!$ times $(0.2)^2$ for some u between $\ln 10$ and $\ln 10 + 2$. Because $f''(u) = e^{-u}$, we know that $0 < f''(u) < f''(a) = 1/10$. Therefore, with $x = \ln 10 + 0.2$,

$$\frac{1}{10} - \frac{0.2}{10} < e^{-x} < \frac{1}{10} - \frac{0.2}{10} + \frac{1}{20}(0.2)^2.$$

Numerically, $0.08 < e^{-(\ln 10 + 0.2)} < 0.082$. The actual value is $0.081873\ldots$.

Here is another example.

Example: Let $f(x) = \cos(x)$, $a = 0$ and $n = 4$. Then $P_4(x) = 1 - \frac{x^2}{2} + \frac{x^4}{24}$. This is also P_5 because $f^{(5)}(0) = 0$. How close is this to the correct value of $\cos x$ at $x = \pi/4$? Because the sixth derivative of cos is $-\cos$, Taylor's theorem says

$$\cos(\pi/4) - P_4(\pi/4) = c(\pi/4)^6$$

where $c = -\cos u/6!$ for some $u \in [0, \pi/4]$. The maximum value of $-\cos$ on $[0, \pi/4]$ is $-\sqrt{1/2}$ and the minimum value is -1, therefore

$$-\frac{1}{720}\left(\frac{\pi}{4}\right)^6 \leq \cos(\pi/4) - P_4(\pi/4) \leq -\frac{1}{720\sqrt{2}}\left(\frac{\pi}{4}\right)^6.$$

For bounds one can compute mentally, we can use the fact that $\pi/4$ is a little less than 1 to get

$$-\frac{1}{720} \leq \cos(\pi/4) - P_4(\pi/4) \leq 0$$

to see that $P_4(\pi/4)$ overestimates $\cos(\pi/4)$ but not by more than $1/720$ which is a little over 0.001.

7 Infinite series

This topic is addressed in Chapter 10 of the textbook, which has ten sections and gets about four weeks in Math 104. That's because it is an important foundation for differential equations, Bessel functions, and Fourier analysis. In Math 110, however, we need this topic for only two reasons, which can be covered in one week. (1) There are some useful infinite sums. In Unit 3 you already saw the most important example of this, the infinite geometric series, which is used for modeling the total lifetime value of an asset or debt. (2) Understanding the Taylor series will help to make sense of the material you just learned concerning Taylor polynomials. For example, is e^x really equal to the infinite sum $1 + x + x^2/2! + x^3/3! + \cdots$?

7.1 Convergence of series: integral test and alternating series

As mentioned before, we are not going to cover the more than a dozen variants of theorems about when infinite series converge. You can get by with just a few methods: comparing to an integral, comparing to a geometric series, and using sign alternation. FYI, the definition of convergence was stated at the end of Section 3.1.

The integral test

Here's an example: Does $\sum_{n=1}^{\infty} n^{-2}$ converge? It's hard to tell from summing the first few terms $1 + 1/4 + 1/9 + \cdots$. This sum should be very similar to $\int_1^{\infty} x^{-2}\, dx$. Does this improper integral converge? Yes. How can we be sure the sum behaves like the integral? We have to somehow compare the sum and the integral. This will be your first in-class problem.

We can generalize this into a theorem, which may be found in Section 10.3 of the textbook on page 594. We will discuss the nuances of the theorem in class.

> **Integral test:** Let $\{a_n\}$ be a sequence of positive terms. Suppose that $a_n = f(n)$ where f is a continuous, positive, decreasing function of x for all $x \geq N$ (N a positive integer). Then the series $\sum_{n=N}^{\infty} a_n$ and the integral $\int_N^{\infty} f(x)\, dx$ either both converge or both diverge.

Alternating series

If the terms $\{a_n\}$ alternate in sign and decrease in magnitude with a limit of zero then $\sum_{n=1}^{\infty} a_n$ must converge. This is more intuitive than it looks. For example, the series

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$$

fails to converge (integral test) but its alternating version

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{\sqrt{n}}$$

converges. To see why, let's write out the series in long form so it is not obscured by notation.

$$1 - \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{4}} + \frac{1}{\sqrt{5}} - \frac{1}{\sqrt{6}} + \cdots$$

In decimal approximations, that's

$$1 - 0.71 + 0.58 - 0.50 + 0.45 - 0.41 + \cdots .$$

Do you see why the partial sums converge? The partial sums are $1, 0.29, 0.87, 0.37, 0.82, 0.41, \ldots$. These alternate down, up, down up,... but notice that after each up the partial sum is not as it was before, and after each down is it not as low as it was before. That means that

> Each partial sum ending in a positive term is an upper bound for the infinite sum;
>
> Each partial sum ending in a negative term is a lower bound for the infinite sum.

This it useful: bounds for alternating series are easy! But also it should make it clear why the theorem is true: you have a bunch of upper and lower bounds getting closer to each other, so the squeeze the series to a limit.

7.2 Convergence of series: ratio and root tests

Let's start with a triviality: if the terms of a series do not go to zero then the series can't possibly converge. Believe it or not, this gets its own theorem, test and example on pages 588–589 in Section 10.2 of the textbook. If the terms do go to zero, then the series still may not converge if the terms get small too slowly. For example, $\sum_{n=1}^{\infty} 1/\sqrt{n}$ does not converge (use the integral test). One case where we know the terms get small fast enough is when they decrease like a geometric series. If $r < 1$ and $a_n \leq Kr^n$ for any constant K, then $\sum_{m=1}^{\infty} a_n$ will converge. (Again, note, for convergence issues, mutliplying by a constant doesn't affect anything.)

It's not always easy to test whether $a_n \leq Kr^n$ and it's not necessary either. If $\sqrt[n]{a_n}$ has a limit, $r < 1$, then a_n is enough like r^n to guarantee convergence. The following statement of this is simplified somewhat from the form in Section 10.5, page 608.

Root test: Suppose $\lim_{n\to\infty} \sqrt[n]{|a_n|} = r < 1$. Then $\sum_{n=1}^{\infty} a_n$ converges.

If you're having trouble computing the limit of $|a_n|^{1/n}$ you can always try the ratio test.

Ratio test: Suppose $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = r < 1$. Then $|a_n|^{1/n} \to r$ and hence the series converges.

In both cases, if $r > 1$ then the terms get big, hence the series can't possibly converge, but if $r = 1$ you don't know anything. Let's try these tests on some series.

EXAMPLE: Does $\sum_{n=1}^{\infty} \frac{n}{2^n}$ converge?

ROOT TEST:

$$\begin{aligned}
\lim_{n\to\infty} \left(\frac{n}{2^n}\right)^{1/n} &= \frac{\lim_{n\to\infty} n^{1/n}}{\lim_{n\to\infty} (2^n)^{1/n}} \\
&= \frac{1}{2}\,.
\end{aligned}$$

Because $1/2 < 1$ we conclude that this series converges.

Ratio test: This is a little easier because it does not require evaluating $\lim_{n\to\infty} n^{1/n}$.

$$\frac{a_{n+1}}{a_n} = \frac{n+1}{n} \cdot \frac{1}{2}$$

which has the obvious limit $\frac{1}{2}$. Again we conclude that $\displaystyle\sum_{n=1}^{\infty} \frac{n}{2^n}$ converges.

Example: Does $\displaystyle\sum_{n=1}^{\infty} \frac{5^n}{n!}$ converge?

Ratio test: Let $a_n = 5^n/n!$. Then $a_{n+1}/a_n = 5/(n+1)$. Clearly $\lim_{n\to\infty} 5/(n+1) = 0$, so the ratio test tells us that this series goes to zero plenty fast for convergence to occur: all that was needed was a limit less than 1, and we managed to get 0.

7.3 Power series

One very important class of infinite series are the **power series** which are series that has a variable x occurring in a very specific way: it has the form $\displaystyle\sum_{n=1}^{\infty} a_n x^n$. If you plug in a real number for x then you get a series that you can try to sum. As it is, it is a function of x, except that for some values of x it may be undefined.

Let me say that again: THIS SERIES IS A FUNCTION OF x. Think this over till it makes sense.

If you can evaluate $\lim_{n\to\infty} |a_{n+1}/a_n|$ and it is equal to a real number r, then $\lim_{n=\infty} |a_{n+1}x^{n+1}/(a_n x^n)|$ will equal $r \cdot |x|$. So you'll get convergence if $|rx| < 1$ and not when $|rx| > 1$. When $|rx| = 1$, you won't know without further examination.

Similarly if you can evaluate $\lim_{n\to\infty} |a_n|^{1/n}$ and you get r, again $\lim_{n\to\infty} |a_n x^n|^{1/n}$ will be $r|x|$. Again the series will converge when $|x| < 1/r$ and diverge when $|x| > 1/r$.

So that's really all you need to know about power series. Here are some examples.

Example: For which x does the series $\displaystyle\sum_{n=1}^{\infty} \left(\frac{2}{3}\right)^n x^n$ converge? The coefficient $(2/3)^n$ is an easy candidate for either the ratio or the root test, resulting in a limt of $r = 2/3$. Therefore the sum converges when $(2/3)|x| < 1$, that is, $|x| < 3/2$ and diverges when $(2/3)|x| > 1$, that is, $x > 3/2$. What about when $|x| = 3/2$? We get no information from the ratio or root tests. The series, for $x = 3/2$, is $\sum_{n=1}^{\infty} 1$, while the series for $x = -3/2$ is $\sum_{n=1}^{\infty}(-1)^n$. In neither case does the term go to zero, so in neither case does the series converge.

Example: For which x does the series $\displaystyle\sum_{n=1}^{\infty} \frac{x^n}{n!}$ converge? Taking $a_n = 1/n!$, we see that $a_{n+1}/a_n \to 0$, so the limiting ratio (and therefore root) is 0. This means we get convergence if $0 \cdot |x| < 1$, in other words, for all real x. Another way of saying this is that factorials grow way faster than any power, or equivalently $x^n = o(n!)$ for any x, which means the terms of the series go to zero quite rapidly.

Example: For which x does the series $\displaystyle\sum_{n=1}^{\infty} (-1)^{n+1}\frac{x^n}{n}$ converge? Because $|a_n| = 1/n$, both ratio and root test result in a limit of $r = 1$. For example, $a_{n+1}/a_n = n/(n+1)$

which clearly converges to 1. Therefore the series converges for $|x| < 1$ and diverges for $|x| > 1$. When $x = 1$ we get the famous **harmonic series** $1+1/2+1/3+1/4+\cdots$. This is one you should commit to memory if you haven't already: by the integral test, because $\int_1^\infty (1/x)\,dx$ diverges, so does the harmonic series. When $x = -1$ we get the **alternating harmonic series** $1 - 1/2 + 1/3 = 1/4 + \cdots$. This converges by the alternating series test, so we see that the power series $\sum_{n=1}^\infty x^n/n$ converges exactly when $x \in [-1, 1)$.

Taylor series

The Taylor series is just the Taylor polynomial with $n = \infty$.

Definition of Taylor series: Let a be any real number and let f be a function that is *smooth* at the point a, meaning it can be differentiated infinitely often. The **Taylor series** for f about the point a is the series

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k \,.$$

For any particular value of x this series may or may not converge. The Taylor series with $a = 0$ is called the **MacLaurin series**:

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k \,.$$

EXAMPLE: What it the Taylor series for the function $f(x) = 1/(1-2x)$ at $x = 0$ and for which values of x does it converge? The easiest way to compute this is by composition. Remember computing the Taylor polynomials for $1/(1-x)$ and getting $P_n(x) = 1+x+x^2+\cdots+x^n$? Evidently the Taylor series for $1/(1-x)$ is $1+x+x^2+\cdots$. Substituting $2x$ for x we find that the Taylor series for $1/(1-2x)$ is

$$1 + (2x) + (2x)^2 + (2x)^3 + \cdots \quad = \quad 1 + 2x + 4x^2 + 8x^3 + \cdots \,.$$

This converges when $|2x| < 1$, thus $|x| < 1/2$. If you graph it, you will see why the series might have problems converging beyond that.

8 Introduction to differential equations

8.1 Modeling with differential equations

A differential equation is an equation, involving derivatives, in which the quantity that you must solve for in order to make the equation true is an unknown *function*. A classical example is Malthusian population growth. This simple population growth model (due to Malthus) postulates that the (instantaneous) rate of growth of a population is proportional to its present size. If we let $A(t)$ denote the size of the population at time t, then the equation representing this is

$$\frac{dA}{dt} = kA(t)\,. \tag{8.1}$$

Here k is a constant of proportionality. It is important that you understand its units! To make the equation work out, k must have units of inverse time. The value $k = 0.03\text{years}^{-1}$ for example, would mean that if you measure time in years, and the present population is one million, then the instantaneous growth rate would be 30,000 people per year.

The *solution* to a differential equation such as (8.1) is any function $A(t)$ that makes the equation true. Typically there will be more than one such equation. For example, the general solution to $A'(t) = 0.03A(t)$ is $Ce^{0.03t}$, where C is any real number. We are going to postpone until Unit 9 the business of to find nice solutions such as this one. In Unit 8 we will concentrate instead on understanding pretty much everything else: how to set up a differential equation, what it means, and what the solution will look like qualitatively.

Verifying that you have found a solution is a lot easier than finding a solution. To check that Ce^{kt} solves (8.1), just evaluate both sides when $A(t) = Ce^{kt}$. The left side is the derivative of $A(t)$ which is Cke^{kt}. The right side is k times $A(t)$ which is Cke^{kt}. They match – whoopee!

The reason you might expect there to be many solutions to an equation such as (8.1) is that it is an equation of evolution. Once you know where you start, everything else should be deterimined, but there is nothing in the equation that tells you where you start. A differential equation together with a value at a certain time is called an *initial value problem*. For example, $A'(t) = 0.03t;\ A(0) = 1,000,000$ is an initial value problem.

Standard form for first-order differential equations

A differential equation could be abitrarily complicated. The equation

$$f(x)\sqrt[3]{1+\left(\frac{df}{dx}\right)^2} - \ln f(x+a) + \exp\left(\frac{d^2 f}{dx^2}\right) = \arctan(x+f(x))$$

is a differential equation but way too complicated for us to have any hope of figuring out what functions f satisfy it. Note the appearance of a second derivative, the square of the first derivative, a big messy cube root and the appearance of the unknown function f as the argument of the arctangent. We will stick to a much simpler class of differential equations, called **first order differential equations in standard form**. This is the form

$$\frac{dy}{dx} = F(x,y)\,. \tag{8.2}$$

Be sure you understand what this means. The unknown function in this case is the function $y(x)$. We call y the "dependent variable" and x the "independent variable". The function F is an abstraction representing that the right-hand side is some function of x and y. Here are some examples:

$$\begin{aligned}
\frac{dy}{dx} &= x^{-2} \\
\frac{dy}{dx} &= ky \\
\frac{dy}{dx} &= x - y \\
\frac{dy}{dx} &= \sqrt[3]{y + e^x}
\end{aligned}$$

Even though you don't yet know much about differential equations, there is a lot you can say looking at these examples. (*i*) It is possible that $F(x,y)$ will be a function of just x, as in the first equation. This means that $y(x)$ is just the integral of this function. So you can already solve this one: it is $y(x) = -1/x + C$. (*ii*) It is possible that $F(x,y)$ will be a function of just y, as is the case in the second equation. In that case it's not so obvious how to solve it, but you actually already know the solution to this particular equation because it is just (8.1). (*iii*) In general, a first order equation in standard form can be simple, like the third one, or complicated, like the fourth. The simple ones are usually exactly solvable (the third one will be solved in Unit 9.3) and the more complicated ones are not. The fourth equation, while not exactly solvable will still yield plenty of information; this is what Unit 8 is mostly about.

A point of notation: should we use y' or dy/dx? Both mean the same thing, but dy/dx is clearer because it tells you which is the dependent variable. If you wrote $y' = -ce^{tx}$ it would be unclear whether t or x was the independent variable (or c could be too, but we never choose c for a variable name because it sounds too much like it should be a constant). We will use both notations, as both are common in real life. One more point: when we want to emphasize that the unknown fariable is a function, we sometimes use a name like f or g instead of y. For example, $f' = -xf$ is a differential equation (it is understood that the indpendent variable is x). The most common independent variable names are x and t, with t usually chosen when it represents time.

Integral equations

Certain equations with integrals in them can be made into differential equations by differentiating both sides (this uses the Fundamental Theorem of Caculus). For example the integral equation

$$f(t) = 12 - \int_5^t 3f(s)\,ds$$

can be differentiated with respect to t to obtain

$$f'(t) = -3f(t)\,.$$

The integral equation has only one solution but this differential equation has many. This means that there was initial value data in the integral equation that we forgot to include in the differential equation. Can you spot it? Really we should have translated the integral equation into the initial value problem:

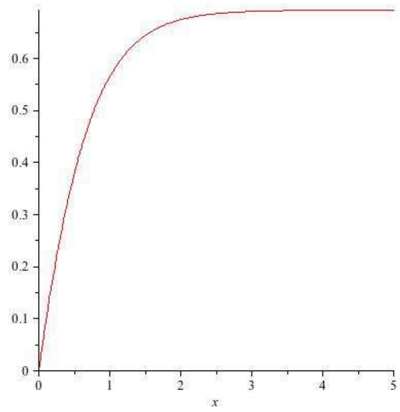
$$f'(t) = -3f(t) \;\;;\;\; f(5) = 12\,.$$

8.2 Slope fields

Slope field drawings are a way to enable you to sketch solutions to differential equations. The equation $dy/dx = f(x, y)$ tells you what the slope of the graph of y should be at any point (x, y), if indeed that point is on the graph. We make a grid of points and through every (x, y) in the grid we put a little line segment of slope $f(x, y)$. We then try to sketch solutions that are always tangent to the line segments, following them as they change direction. Pages 538–539 of the text do a good job explaining this. PLEASE READ THESE! We will then spend a day practicing.

Slope fields are a *qualitative* approach to understanding the solution to a differential equation, meaning that you get information about the nature of the solution even when you can't find the exact solutions. Here's an example.

Why can we tell that the solution to $y' = 2 - e^y$ should approach a limit of $\ln 2$? It's because when $y(x) < \ln 2$ then $2 - e^y$ is positive and the function threrefore increases while when $y(x) > \ln 2$ then $2 - e^y$ is negative and the function therefore decreases. It seems clear from this that if the function begins below $\ln 2$ it will steadily increase but at a lesser and lesser rate and never get above $\ln 2$, while if the function begins above $\ln 2$ then it will steadily decrease but at a lesser and lesser rate and never get below $\ln 2$. We therefore have a very good idea what this function looks like without ever solving the equation:



8.3 Euler iteration

Euler's method, or "Euler iteration" is a way of finding a numerical approximation for the solution of an initial value problem at some later time. In other words, for the equation $y' = f(y,t); y(t_0) = y_0$, you can compute an approximation to $y(t_1)$ when t_1 is any time greater than t_0.

The idea behind Euler iteration is that you follow the slope field for a small amount of time Δt, which is fixed at some value such as 0.5 or 0.1. Let $t_1 = t_0 + \Delta t$ be the new time and let y_1 be the approximation you get by followingthe slope field fo time Δt. In other words, $y_1 = y_0 + (\Delta t) f(y_0, t_0)$. The slope at the point (t_1, y_1) will in general be different. Follow that slope for time Δt, and repeat.

I don't have a lot to add to what's in the textbook on Page 539–541. Euler's method is important because it gives you an in-principle understanding of what a solution should be like, whether or not you can produce an analytic solution. This is important for your understanding even if you rarely use Euler's method in practice.

Different notions of solution

Our last order of business in this section is some philosophy. You need to understand what is meant by a solution to a differential equation. The simplest differential equation is of the form $y' = f(x)$, in other words, the right-hand side does not depend on y. You already know how to solve this: $y(x) = \int f(x)\, dx$. But wait, what if it's something you can't integrate? An example of this would be $dy/dx = e^{x^2}$. We could write a solution like $y(x) = y(0) + \int_0^x e^{t^2}\, dt$, but is this really a solution? The answer is yes. Here's why.

Euler's method allows you to approximate values of the independent variable. For example, given $y' = f(x, y)$ and $y(0) = 5$ we could use Euler's method to evaluate $y(2)$. What you need to understand is that yes we can do it but it's tedious and not all that accurate unless you use a miniscule step size. By contrast, using Riemann sums to estimate $\int_0^2 e^{t^2}\, dt$ is a piece of cake. Keep in mind the relative difference in difficulty between Riemann sums and Euler's method as we discuss three levels of possible solution to a differential equation.

1. If you can find a solution $y = f(x)$ where f has an explicit formula then that is obviously the best. Your calculator or computer (or maybe even your phone) can evaluate this, and typically you have other information associated with f such as how fast it grows, whether it has asymptotes, and so forth.

2. Next best is if you can write a formula for y that involves functions without nice names, defined as integrals of other functions. You already realize that many simple looking functions such as e^{x^2} and $\ln(x)/(1+x)$ have no simple anti-derivative. The differential equation $y' = e^{x^2}$ is trivial from a differential equations point of view (it is in the form $y' = f(x)$ which we discussed above) but still we can do no better than to write the solution as $y = \int e^{x^2}\,dx$. This is perfectly acceptable and counts as solving the equation.

3. Lastly, for the majority of equations, we can't write a solution even if allowed to use integrals of functions. In this case the best we can do is to numerically approximate particular values and to give limiting information or orders of growth for y. For example, if $y' = 2 - e^y$ then $\lim_{x\to\infty} y(x) = \ln 2$.

One last thing that Euler iteration does for us is to convince us that an initial value problem should have a solution. After all, if you look at an equation with functions and derivatives, there is no reason to believe that there is a function satisfying the equation. But Euler iteration shows you that there has to be. Just do Euler iteration and make the steps smaller and smaller; in the limit it will produce a function satisfying the differential equation. This is the basis for a theorem. The theorem is not officially part of this course but you might be interested to know what it says.

> **Theorem:** Let $f(x, y)$ be a continuous function. Then the initial value problem $y' = f(x, y), \quad y(x_0) = y_0$ has a unique solution, at least for a small amount of time (after that it might become discontinuous). This solution can be obtained by taking the limit of what you get from Euler iteration as the step sizes go to zero.

9 Exact solutions to differential equations

This unit covers Sections 7.2 and 9.1–9.2 of the textbook. It concerns mainly techniques of computation. For each of the three class days I will give a short lecture on the technique and you will spend the rest of the class period going through it yourselves.

Exactly solving differential equations is like finding tricky integrals. You have to recognize the equation as a type for which you know a trick, then apply the trick. You will learn precisely two tricks. The first works for a class of equations called **separable equations**. The trick involves getting all the x variables on one side of the equation and the y variables on the other (hence the name "separable"). The second class is the class of linear first order equations. The trick there will be to find a so-called integrating factor. Before learning either of these tricks, we'll spend a day getting familiar with the easiest but single most important differential equation. This one is bother separable and linear.

9.1 $f' = kf$ and exponential trajectories

The single most important differential equation is, as luck would have it, very easy to solve:

$$\frac{dy}{dx} = ky \tag{9.1}$$

where k is a constant. You can solve it by guessing the answer but let's solve it a way that will generalize.

Step 1: Separate. To get the dependent variables on the left and the independent variables on the right, we divide both side by y and multiply both sides by dx:

$$\frac{dy}{y} = k\,dx\,. \tag{9.2}$$

If you are worried about whether it's OK to multiply dy/dx by dx, you're right to be concerned, because a dy without a $/dx$ is meaningless, but it works anyway and we'll show you why later.

Step 2: Integrate. The integral of dy/y is $\ln|y|$. What's the integral of k. You might think it's $1/2k^2$ but it's not. Pay attention to the variable of integration, which is dx.

The integral of $k\,dx$ is just kx. So now (don't forget the constant) we have

$$\ln |y| = kx + C\,. \tag{9.3}$$

Step 3: exponentiate both sides to get

$$|y| = e^{kx+C} = e^C e^{kx}\,. \tag{9.4}$$

Step 4: The constant e^C can be any *positive* real number. If the absolute value of y can be any positive multiple of e^{kx} that's the same as saying y can be any multiple of e^{kx} positive or negative. Call this multiple C_1. We write the solution in its final form:

$$y = C_1 e^{kx} \tag{9.5}$$

where C_1 can be ay real number.

Note: A lot of people like to call the new constant C_1 the same name as the old constant and write $y(x) = Ce^{kx}$, where C is any real number. This is a correct solution, but we don't want you changing the value of C midstream if it leads to writing incorrect equations such as $e^C e^{kx} = Ce^{kx}$.

In an application, the independent variable will be expressed in some natural unit, often time, and the function variable will have another unti such as money, volume, total quantity of something, etc. The units of dy/dx are y-units divided by x-units, so in the equation $dy/dx = ky$, the units of the constant k must be in units of "reciprocal x". For example, if x is in seconds then k is in $(\text{sec})^{-1}$: the name for this unit is Hertz, abbreviated Hz. In the solution $y = Ce^{kx}$, notice that the exponent is unitless (as I have previous claimed must be true of exponents). It is good to be aware of this, as it is one easy way to doublecheck your work.

The meaning of these equations in applications is that the rate of change is proporitional to the present quantity, or to the difference between the present quantity and some limiting value. Therefore we should think of k as a **relative rate of change**, that is a rate expressed as a fraction of the whole. One example is an interest rate. Interest, even though it produces dollars, is measured in units of inverse time because the dollars cancel: each dollar begets a similar number of future dollars. The number of dollars produced in a year by a single dollar is equal to the number of pennies produced in a year by a single penny or the number of gigabucks produced in a year by a single gigabuck.

Applications

We will look at a number of applications. There is a fundamental difference in the behavior depending on whether k is positive or negative. When k is positive, e^{kx} grows (in fact very rapidly). When k is negative, e^{kx} shrinks.

One quantity that tends to grow exponentially is wealth. Wealth can be negative if it's a debt, but both debts and assets tend to grow rather than shrink. Another is population. A characteristic of applications in which $f' = kf$ with $k > 0$ is that each unit of whatever quantity is growing contributes to the growth independently of each other unit. So for example, if you put two chunks of money in two accounts at the same interest rate, it's just like putting in one chunk that's the sum of the original two. This is reflected in the fact that the growth rate is a proportion per time, not an absolute amount per time.

Let's look at $y' = -ky$ more closely. It approaches zero. Very closely related is when a quantity comes to equilibrium at some value other than zero. For example, suppose an object at temperature A is placed in a large bath at temperature B. The temperature of the object approaches the temperature of the bath at a rate proportional to the difference in tempearatures. Mathematically, if the temperature as a function of time is denoted by $y(t)$, we have

$$\frac{dy}{dt} = k(B - y)\,.$$

Let's see two ways of solving this. One is using the same method as before. We get the chain of equations:

$$\begin{aligned}
\frac{dy}{B-y} &= k\,dt \\
-\ln|B-y| &= kt + C \\
|B-y| &= e^{-C-kt} \\
B - y &= \pm e^{-C}e^{-kt} = C_1e^{-kt} \\
y &= B - C_1e^{-kt}
\end{aligned}$$

where C_1 is any real number.

What is this saying qualitatively? If the dependent variable is approaching a target, B, at a rate proportional to the distance from the target, then the value at time t will be the equilibrium value B plus an offest that decreases exponentially. Many physical systems behave this way (thermal equilibria, radioactive decay, resistor-capacitor networks) but also systems in social science where there is negative feedback (population approaching a natural limit, price corrections after an economic shock, etc.).

The second way to solve this is to move upward by B: algebraically, replace y by $y - B$. We will discuss this in class.

Continuous versus annualized rates

Interest is fundamentally a continuous time phenomenon, especially when the amounts involved are so large that the interest is substantial even in a minute or a few seconds (think National Debt). Consumers, however, are barely able to handle simple interest and haven't a clue about continous time interest. This has led to regulation where interest rates must be quoted in Annualized Percentage Yield (APY) as well as a simple growth rate.

To see how this works, suppose an asset grows at the rate of 6% per year. If $A(t)$ is value at time t, this means that $A'(t) = 0.06A(t)$ when t is measured in years; the units of the 0.06 are inverse years. After one year, an amount A_0 will grow to $A_0e^{0.06}$. That means the gain was $A_0(e^{0.06} - 1)$. Because $e^{0.06} \approx 1.061837$, this means that the percentage growth of the asset in one year was roughly 6.1837%. In other words:

Continuous interest rate of 6% leads to annualized interest rate of 6.1837%.

We can do this for any rate. Let r be the continuous rate (in the above example 0.06) and let A be the annualized rate (in the above example 0.061837). Then r and A are related by the equations:

$$A = e^r - 1 \ \ ; \ \ r = \ln(1 + A)\,.$$

If you write these as percentages, you have to remember to multiply and divide by 100 at the apropriate places:

$$A = 100(e^{r/100} - 1) \ \ ; \ \ r = 100\ln(1 + A/100)\,.$$

You can also read about this in Hughes-Hallett, page 571.

9.2 Separable equations

The book teaches this clearly and succinctly in half a page starting at the top of page 432, with two worked examples immediately following. I probably don't need to repeat it for you here, but to summarize very briefly, the steps are:

1. Recognize the equation as having the form
$$\frac{dy}{dx} = \frac{g(x)}{h(y)}$$

2. "Cross-multiply" to get
$$h(y)\, dy = g(x)\, dx$$

3. Integrate both sides to get the equation
$$H(y) = G(x) + C$$
where H is an anti-derivative of h and G is an anti-derivative of g.

4. Solve for y by applying the inverse function to H:
$$y(x) = H^{-1}(G(x) + C)\,.$$

If it bothers you that the equation $h(y)\, dy = g(x)\, dx$ is not a real equation because dy and dx aren't actual numbers and have meaning only as symbols in integrals such as $\int g(x)\, dx$, then you should read the justification in the middle of page 432.

EXAMPLE: $y' = x + xy$. You can write this as $y' = x(1 + y)$ so it is of the right form with $g(x) = x$ and $h(y) = 1/(1 + y)$. Please see the discussion in the book about the form $g(x)H(y)$ being the same as the form $g(x)/h(y)$ with $h(y) = 1/H(y)$ but don't be confused: the H in that discussion is NOT the H that's the anti-derivative of h (I have no idea why they use H instead of some other letter!). Continuing, we write
$$\frac{dy}{1 + y} = x\, dx$$
and integrate both sides leading to
$$\ln|1 + y| = \frac{x^2}{2} + C\,.$$

In other words, $H(y) = \ln(1+y)$ and $G(x) = x^2/2$. The calculus is now done. You can use algebra to write the equation in a more understandable form. To isolate y, exponentiate both sides and subtract one. Here is the sequence of equations.

$$\begin{aligned}
\ln|1+y| &= \frac{x^2}{2} + C \\
|1+y| &= ce^{x^2/2} \text{ where } c = e^C \text{ is any positive constant} \\
1+y &= ce^{x^2/2} \text{ where } c = \pm e^C \text{ is any constant} \\
y &= -1 + ce^{x^2/2}\,.
\end{aligned}$$

That's it! Practice it. Learn it. Your brain is built to do procedures like this and you will probably find that you learn this one with relative ease. The rest of what I have to say about separable equations concerns some qualitative aspects of their solutions.

9.3 Integrating factors and first order linear equations

The exposition in Section 9.2 is very clear. The only thing I have to add is an explanation of how to find the right constants to solve an initial value problem in the case that you can't xplicitly integrate one of the functions you need to integrate. Begin with the equation

$$y' + P(x)y = Q(x)$$

and suppose that P has no closed form integral. Letting $v(x) = \exp(\int P(x)\,dx)$ we end up (see page 545) with

$$y = \frac{1}{v(x)} \int v(x)Q(x)\,dx\,. \qquad (*)$$

Adding an arbitrary constant c to $\int P(X)\,dx$ in (*) does not change the answer at all: it multiplies the integral for $v(x)$ by $C = e^c$, reducing the outside $1/v(x)$ by a factor of C but increasing the integrand by the same factor. However, an additive constant in the integral of $v(x)Q(x)$ produces the general solution

$$y(x) = \frac{1}{v(x)} \left(\int v(x)Q(x)\,dx + C \right)\,.$$

These represent different solutions as C varies. Suppose you have an initial condition $y(x_0) = y_0$. The computation is easiest if you make the definite integral start at x_0.

This gives

$$y = \frac{1}{v(x)}\left(\int_{x_0}^{x} v(t)Q(t)\,dt + C\right).$$

The integral vanishes at x_0 leaving $C/v(x_0)$. Therefore, you need to set C equal to $y(x_0)v(x_0)$. To summarize: the soluition to $y' + P(x)y = Q(x)$ with initial condition $y(x_0) = y_0$ is

$$y(x) = \frac{1}{v(x)}\left(\int_{x_0}^{x} v(t)Q(t) \;+\; y(x_0)v(x_0)\right)$$

For those who would like this broken down into steps, here are the steps.

1. Get the equation into the form

$$y' + P(x)y = Q(x)$$

2. Compute the integrating factor

$$v(x) = e^{\int P(x)\,dx}$$

   Note: you do not have to worry about the $+C$ in this step; any choice works.

3. Multiply through by $v(x)$ and integrate. On the left-hand side you don't have to do the integral because you know it is going to be $v(x)y(x)$. The equation is now

$$v(x)y(x) = \int v(x)Q(x)\,dx\,.$$

   Note: the integral on the right-hand side may or may not be do-able, but in either case, this time you need to include the $+C$.

4. Divde by $v(x)$ and you're done.

Part IV

Multivariable calculus

Before we tackle the very large subject of calculus of functions of several variables, you should know the applications that motivate this topic. Here is a list of some key applications.

1. Totals of quantities spread out over an area.
2. Probabilities of more than one random variable: what is the probability that a pair of random variables (X, Y) is in a certain set of possible values?
3. Marginal cost.
4. Optimization: if I have a limit on how much I can spend on production and advertising in total, and my profit will be some function $f(p, a)$, then how much should I invest in production and how much in advertising?

When dealing with these sorts of questions, the functions and their notation can start to seem difficult and abstract. Geometric understanding of multi-variable functions will help us think straight when doing word problems and algebraic manipulations.

10 Multivariable functions and integrals

10.1 Plots: surface, contour, intensity

To understand functions of several variables, start by recalling the ways in which you understand a function f of one variable.

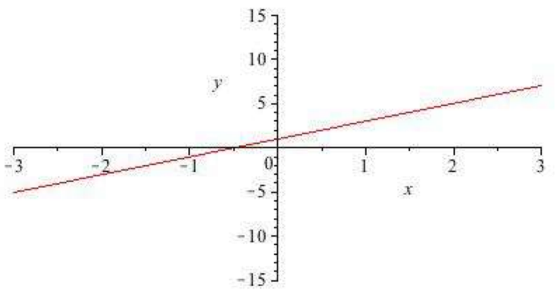
(i) As a rule, e.g., "double and add 1"

(ii) As an equation, e.g., $f(x) = 2x + 1$

(iii) As a table of values, e.g.,

x	0	1	2	5	20	-95	π
$f(x)$	1	3	5	11	41	-189	$2\pi + 1$

(iv) As a graph, e.g.,



Similarly, a function f of two variables is a way of associating to any pair of values for x and y (two real numbers) a real number $f(x, y)$. The same options apply for understanding f.

(i) We can give the rule if it is easily stated, e.g., "multiply the two inputs."

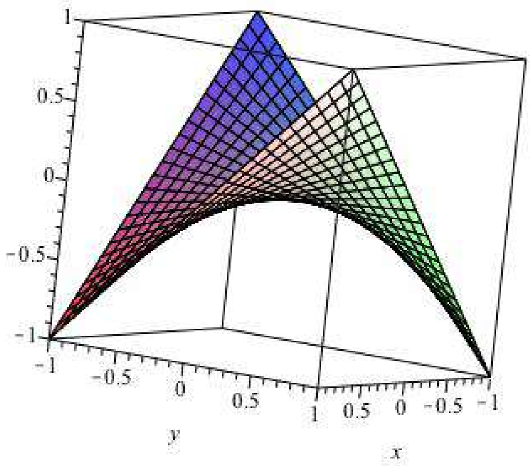
(ii) We could give an equation, such as $f(x, y) = xy$.

(iii) We could make a table, e.g.,

x	1	1	1	2	2
y	0	1	5	0	π
$f(x, y)$	0	1	5	0	2π

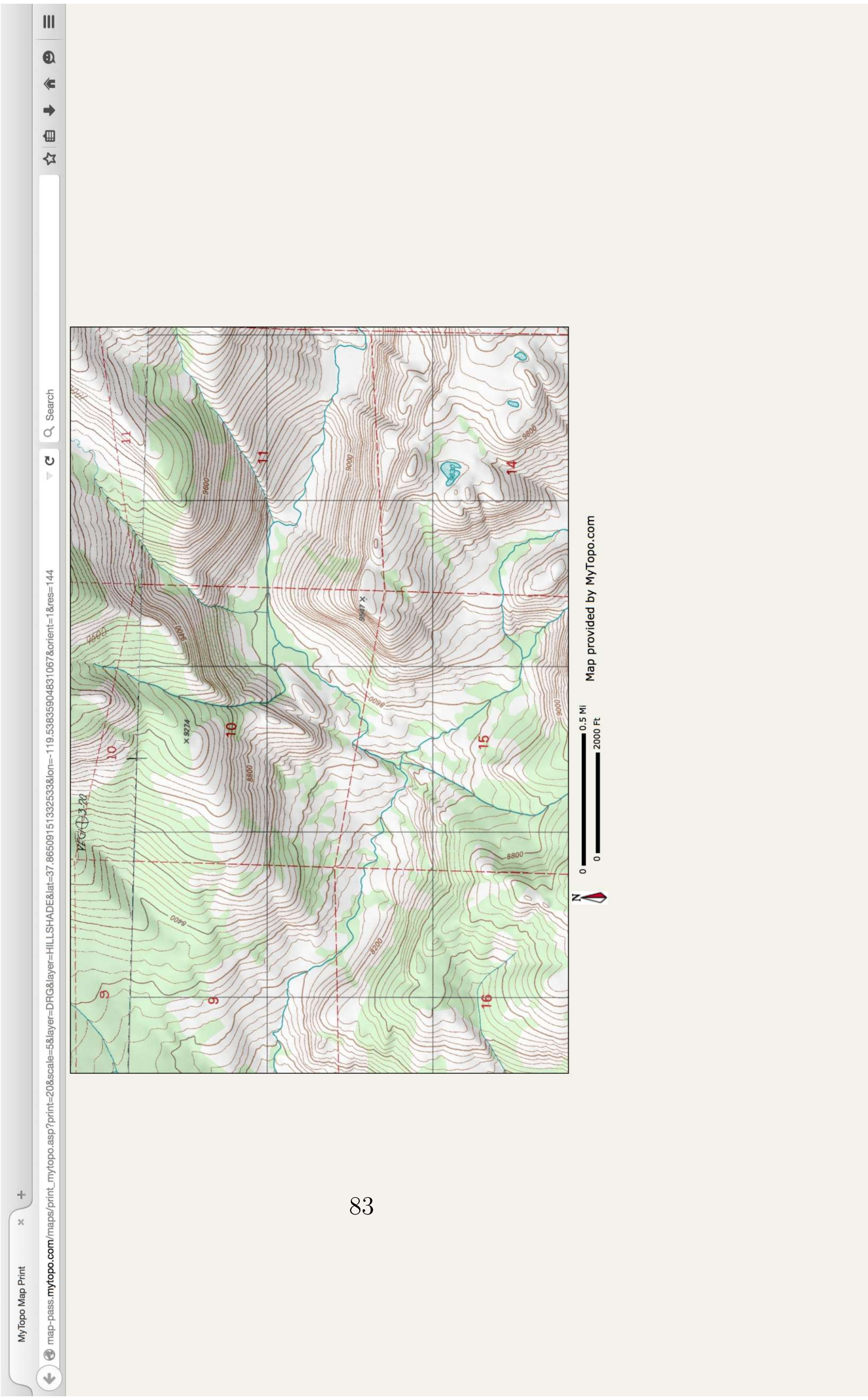
(iv) One might graph f.

You can think of a function of two variables as having two real inputs x and y or as having one input that is a pair (x, y). The second way makes the domain of the function into (some subset of) the xy-plane. For more on how to figure out exactly what subset forms the domain, look at the first few pages of Section 14.1 of the textbook. We won't focus on that, but we will use geometry to understand f via its various visual depictions. There most common way to make a graph of f is to plot the three-dimensional surface $z = f(x, y)$ as in the following figure.

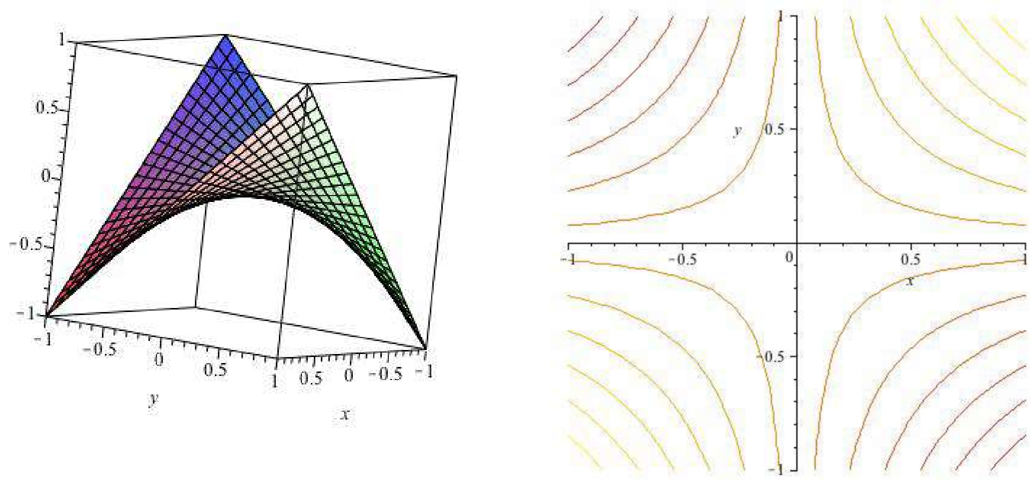


Another way is to plot the **level curves**. To do this, you have to figure out which points (x, y) share the same f-value, say zero, and draw a curve indicating that set. Then, draw the curve indicating another nearby value such as $1/2, 1, -1$, etc. This is shown on the right of the figure above. The book explains this on page 797. The convention when drawing level curves is to pick some fixed increment, such as every $1/2$ or every 100, and draw the level curves corresponding to these regular intervals.

The US Geological Service produces a series of maps drawn this way. These are contour plots of $f(x, y)$, where f is the elevation and x and y are distance east and distance north of the center of a quadrangle.


MyTopo Map Print
map-pass.mytopo.com/maps/print_mytopo.asp?print=20&scale=5&layer=DRG&layer=HILLSHADE&lat=37.86509151332533&lon=-119.53835904831067&orient=1&res=144
Search
Map provided by MyTopo.com
0.5 Mi
2000 Ft

The elevation example is very important even if you don't care about hiking. This is because the traditional way to plot f is to plot the surface $z = f(x,y)$, which means that our brains are primed to accept $f(x,y)$ as an elevation at the point with coordinates (x,y). However, this is far from the only use of contour plotting. The most important application of this is when $f(x,y)$ is profit or some other kind of a utility function (e.g., the level of satisfaction when you have x dollars in the bank and a car that costs y dollars). The contour plot of f shows the **indifference curves**. Later we can use this interpretation of contour plots along with some calculus to compute optimal allocations. The next figure shows the contour plot for $f(x,y) = xy$ along with the height plot $z = f(x,y)$ that you already saw for this function.



All we are doing in this first section is getting used to functions of more than one variable and their visual depictions. We're almost done, except that we haven't talked about functions of three or more variables. We don't have four dimensions handy, so we can't graph $z = f(x_1, x_2, x_3)$. We can still think of f as a function mapping points in an abstract n-dimensional space to the real numbers, and in the case of exactly three variables, we can make a contour plot which now has contour surfaces in three dimensions; see Figure 14.8 in the book. For now, it suffices to practice going back and forth between the equation for a function of two variables and its visual representations.

10.2 Multivariate integration: rectangular regions

This section is a bit heftier than the previous one because multiple integrals are really, really important. This is a tricky topic for two reasons. First, students often confuse the *definition* of a double integral with the *computation* of a double integral. I will try to help you keep these straight. Secondly, non-rectangular regions of integration (which are the topic of Section 10.3) require deeper understanding of free and bound variables than you have needed for the calculus you've done so far. Please come to class having read Section 15.1 of the textbook!

(i) Meaning

Let R be a region in the xy-plane and let $f(x, y)$ be a function. The notation $\int_R f(x, y)\, dA$ is read as "the double integral of f over the region R" and defined as follows (I am parphrasing what is on page 883 of your textbook).

> Divide R into small rectangular regions (ignore for now the fact that these don't quite cover R or sometimes extend a litle beyond R). Multiply the are of each rectangle by f evaluated at some point in that rectange, and add up all of these products. The integral is defined to be the limit of this sum of products as the rectangles get small.

What does this compute? In general it computes the total amount of stuff when f is a density of stuff per unit area. For example, suppose the density of iron ore over a patch of ground is a function $f(x, y)$ that varies due to proximity to some pre-historic lava flow. Then $\int_R f(x, y)\, dA$ will be the total amount of iron ore in the region R. Do you see why? You can get the total by adding up the amount in regions small enough that f doesn't vary significantly; then the amount of ore in the region is roughly the area times f evaluated at any point in the region, so we should expect that adding up these products approximates the total; in the limit, it *is* the total.

Time for a bunch of conceptual remarks!

1. Notice there is now a quantity dA rather than dx or dy. This means, literally, "the teeny amount of area". Starting now, it will be very important to keep track of the infinitesimals.

2. The units of $\int_R f(x,y)\,dA$ are units of f times units of A. The units of A can be area, but more generally, they are whatever unit x represents times whatever unit y represents.

3. Try to see how this is analogous to integrals in one variable. In each case you break up the (interval / region), then in each small part you evaluate f somewhere, use this as a proxy for f throughout the small part, multiply by the (length / area) of the small part, sum and take the limit.

4. You can integrate in three variables. Just chop a 3-D region into subregions, sum their volumes times the value of $f(x,y,z)$ somewhere in the region, and take a limit. In fact, you can do this in any number of variables even though we can't visualize space in dimensions higher than three. In Math 110, we'll stick to two variables.

Here are some more meanings for a double integral.

Volume. If $f(x,y)$ is the height of a surface at the point (x,y), then $\int f(x,y)\,dA$ gives the volume underneath the surface but above the xy-plane. That's because the summands (namely the area of a little region times $f(x,y)$ evaluated at a point in the region) is the volume of a tall skinny rectangular shard, many of which together physically approximate the region. If you can't picture this, you have to have a look at Figure 15.3. Notice the units work: f is height (units of length) and $\int_R f(x,y)\,dA$ is volume, which does indeed have units of length times area.

Area. A special case is when $f(x,y)$ is the constant function 1. Who would have thought that integrating 1 could be at all important? But it is. If you build a surface of height 1 over a region R, then the volume of each shard is the area at the base of the shard and the integral is just the limitin sum of these, namely the total area. Notice the units work: in the example f is unitless, and $\int_R f(x,y)\,dA$ is the area of R, which has units of area.

Averages. By definition, the average of a varying quantity $f(x,y)$ over a region R is the total of f divided by the area of the region:

$$\text{Average of } f \text{ over } R = \frac{\int_R f(x,y)\,dA}{\text{Area of } R}\,.$$

Probability. This application will get its own treatment in Section 10.4.

(ii) **Computing the iterated integral: rectangular regions**

Remember how it worked when you learned integration in one variable? It was defined as the limit of Riemann sums, which intuitively captures the notion of area under a curve. Then there's a theorem saying you can figure out the value of the integral over an interval by computing an antiderivative and subtracting its values at the two endpoints. Similarly, we have already defined the integral conceptually, now we need to say something about using calculus to compute it. A lucky fact: we don't need anything as difficult as the Fundamental Theorem of Calculus like we did for one variable integrals. That's because we assume you already know how to compute single variable integrals and that can be harnesses to compute the double integral. Remember, for now we're sticking to to the case where R is a rectangle.

As the textbook does, we start by assuming R is a rectangle $a \leq x \leq b$ and $c \leq y \leq d$, chopped up so that each little square has width Δx and length Δy. We then add up the little bits in an organized way. First add all the tall skinny rectangles over a given x interval as y varies. In the volume interpretation this gives the volume of the slice of the solid that has width Δx. There is a slice for each x-value in the grid.

Here's the thing. If you fix a value $x = M$, then **you're just computing Δx times the area under the one-variable function** $f(M, y)$. You know how to do that: you integrate $\int_c^d f(M, y)dy$ and multiply by Δx. This integral of course depends on M. Call it $g(M)$. Summing all the slice volumes is the same as integrating $g(M)$ from a to b. We don't have to use the variable M, we can just call it x. So the answer is:

$$\int_R f(x, y)\, dA = \int_a^b g(x)\, dx, \quad \text{where } g(M) = \int_c^d f(M, y)\, dy\,.$$

This is Fubini's Theorem (first form) on page 885 which you practiced computing in the MML problems from Section 15.1. I prefer to put parentheses into the equation given in the book:

$$\int_R f(x, y)\, dA = \int_a^b \left[\int_c^d f(x, y)\, dy \right] dx = \int_c^d \left[\int_a^b f(x, y)\, dx \right] dy\,. \tag{10.1}$$

At this point it would be a good idea to read Examples 1 and 2 in Section 15.1. Also, you should pay attention to free and bound variables. In the so-called inner integral $\int_c^d f(x, y)\, dy$, the variable y is bound, but x is free. In other words, this integral represents a quantity that depends on x (but not y). That's why we can integrate it against dx in the outer integral, to finally get a number.

In Example 1 of Chapter 15 of the textbook, they do the integral two ways (x-direction first versus y-direction first) to show that you get the same answer (that's part of Fubini's Theorem). Sometimes you need to use this to evaluate an integral that appears difficult: write it in the other order and see if it is easier; one of the homework problems is on this technique.

Magic product formula

Suppose your function $f(x, y)$ is of the form $g(x) \cdot h(y)$ and your region of integration is a rectangle $[a, b] \times [c, d]$. Then

$$\int_R f(x,y)\,dA = \left(\int_a^b g(x)\,dx\right) \times \left(\int_c^d h(y)\,dy\right) .$$

Can you see why? It's due to the distributive law. The Riemann sum for the double integral actually factors into the product of two Riemann sums. I'll do this on the blackboard for you because, when written without narration, it just looks like a mess.

One parting word: circling back to the issue of distinguishing the definition from the computation, the left-hand side of (10.1) refers to the definition – a limit of Riemann sums; the two expressions after the equal signs are single variable integrals, computable as antiderivatives. The theorem is asserting that they are all equal.

10.3 Multivariate integration: general regions

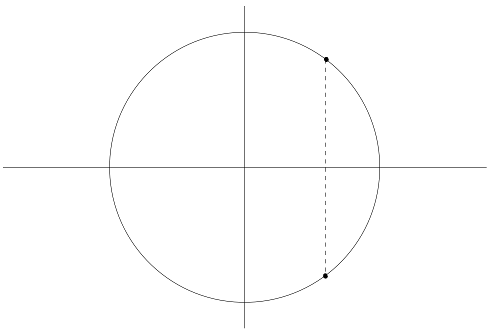
The trickiest thing about learning double integration is when R is not a rectangle. Then, when you cut into slices, the limits of integration will change with each slice. That's OK as long as you can write them as a function of the variable you are not integrating and evaluate properly. There are four examples in the book (Section 15.2), plus I'll give you one more here. But before diving into these, we should review how to write sets of points in the plane.

Writing sets of points in the plane

The notation $\{(x, y) : \text{ blah blah blah}\}$ denotes the set of points in the plane satisfying the condition I have called "blah blah blah". For example, $\{(x, y) : x^2 + y^2 \leq 1\}$ is the unit disk. You will need to become an expert at writing sets of points in a very specific manner: the set of points where x is in some interval $[a, b]$ and y lies between two functions of x, call them g and h. It looks like

$$\{(x, y) : a \leq x \leq b,\ g(x) \leq y \leq h(x)\}\ .$$

Example: can you write the unit disk in this format? For a and b you need the least and greatest x values that appear anywhere in the region. For the unit disk, that's -1 and 1. Then, for each x, you need to figure out the least and greatest y values that can be associated with that x. For the unit disk, the least value is $-\sqrt{1 - x^2}$ and the greatest is $+\sqrt{1 - x^2}$.



The y-value goes from $-\sqrt{1 - x^2}$ to $+\sqrt{1 - x^2}$

So in the end, the unit disk $\{(x,y) : x^2+y^2 \le 1\}$ can be written in our standard form as

$$\{(x,y) : -1 \le x \le 1,\ -\sqrt{1-x^2} \le y \le \sqrt{1-x^2}\}\,.$$

This way of writing it naturally breaks the unit disk into vertical strips where x is held constant and y varies from some least to some greatest value depending on x. I should have said this is "a standard form" not "the standard form" because it is equally useful to break into horizontal strips. These correspond to the format

$$\{(x,y) : c \le y \le d,\ g(y) \le x \le h(y)\}$$

where for each fixed y, the x values range from some minimum to some maximum value depending on y. You will be practicing a lot with these two formats!

Limits of integration for non-rectangular regions

What I am explaining here is Theorem 2 on page 889 of the textbook. When computing $\int_R f(x,y)\,dA$, if you can write R as a region in the form above.

There are three steps. First, specify the region of integration in terms of varying limits of integration. Second, use these as limits of integration. If x goes from z to b while y goes from $g(x)$ to $h(x)$ then the integral will look like $\int_z^b \int_{g(x)}^{h(x)} f(x,y)\,dy\,dx$. Third, carry out the integration with these limits.

Example: Let R be the unit disk and let $f(x,y) = 1$. The possible x-values in R range from -1 to 1. So we put this on the outer integral: $\int_{-1}^1 [\ \cdots\ ]$. Now fix a value of x and figure out what the limits are on y. As we have seen, y goes from $-\sqrt{1-x^2}$ to $\sqrt{1-x^2}$. So now we can write the whole integral as

$$\int_{-1}^1 \left[\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1\,dy\right]\,dx\,.$$

When we do the inner integral we get the antiderivatite y, which we evaluate at the upper and lower limits: $y\,\big|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} = 2\sqrt{1-x^2}$. Finally, we evaluate the outer integral, obtaining $\int_{-1}^1 2\sqrt{1-x^2}\,dx$. This is a tough integral if you do it honestly: integrating by parts and using #18 in the integral table will give you

$$\left(x\sqrt{1-x^2} + \arcsin x\right)\Big|_{-1}^{1}\,.$$

The value of $x\sqrt{1-x^2}$ is zero at both endpoints, so this evaluates to $\arcsin(1) - \arcsin(-1) = \pi/2 - (-\pi/2) = \pi$.

Here are an FAQ about what we just did.

1. When we took the anti-derivative of the constant function 1 why did we get y and not x? Answer: we were integrating in the y-variable at that time.

2. How can you know whether the limits on the inner integral will be functions of y or functions of x? Answer: if you choose vertical strips then the inner integral is dy, the outer integral is dx and the limits on the inner integral can be functions of x but not y.

3. Is it a coincidence that after a complicated computation, the integral came out to be a very simple expression? Answer: No! It's because the integral of 1 over a region gives the area, and the area of a circle is a very simple expression. In fact, if you were asked to do this integral on a test or homework, you should probably not do any calculation and just say it's the area of a circular region with radius 1 and is therefore equal to π.

Switching the order of integration

You have seen how to take a region R and write it in either standard form: horizontal or vertical strips. Sometimes, in order to make an integral do-able, you will want to switch from horizontal strips to vertical strips or vice versa. Starting with one standard form, you convert to a geometric region R, then write that in the other standard form. This allows you to switch between an iterated integral with x in the inside and one with y on the inside.

EXAMPLE: Compute $\displaystyle\int_0^1 \int_y^1 \frac{\sin x}{x}\, dx\, dy$. Unfortunately you can't integrate $\sin x/x$. But wait! The region $\{0 \le y \le 1,\ y \le x \le 1\}$ is triangular and can also be written in vertical strips: $\{0 \le x \le 1,\ 0 \le y \le x\}$. The integral is therefore equal to $\displaystyle\int_0^1 \int_0^x \frac{\sin x}{x}\, dy\, dx$. We can now see that this is equal to

$$\int_0^1 \left(y \frac{\sin x}{x} \right) \Bigg|_0^x dx = \int_0^1 \sin x\, dx = 1 - \cos(1)\,.$$

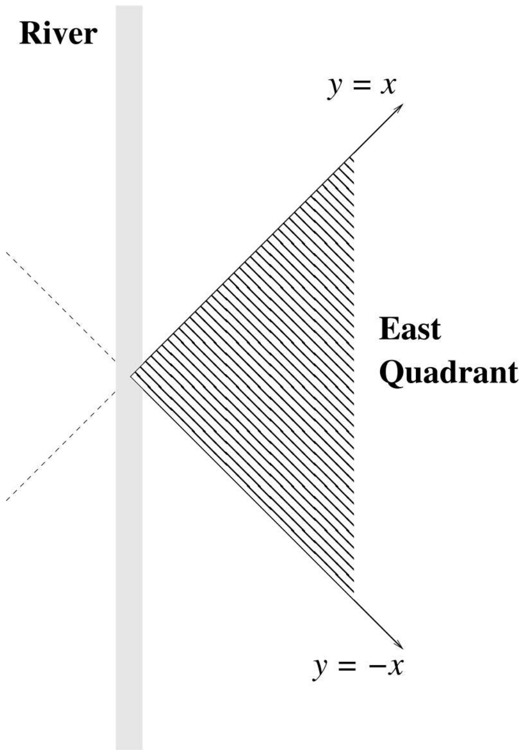
10.4 Applications: spatial totals, averages, probabilities

No new math in this section, just some applications. Two of them are pretty straightforward: integrals to yield total amounts and integrals to compute averages. The third, probability densities in two variables, will involve a couple of new concepts.

Integrals to compute totals

This is essentially just a reminder that the integral of stuff per unit area over an area yields total stuff.

EXAMPLE: The population density east of a river running north-south is $f(x,y) = 6000e^{-x^2}$ people per square mile. The county is divided into quadrants as shown in the figure. Roughly how many people are there in the east quadrant?



Solution: Make coordinates in which the quadrant is the region represented in standard form by the set

$$\{(x, y) : 0 \leq x < \infty, -x \leq y \leq x\}\,.$$

We're not going to re-do the theory of improper integrals in two variables, we'll only deal with cases where you can just plug in ∞ and get the right answer. The region is in standard form, so the total population is given by

$$\int_0^\infty \int_{-x}^{x} 6000 e^{-x^2}\, dy\, dx\,.$$

The inner integral might look tough but it's not (look carefully at which is the variable of integration):

$$\int_{-x}^{x} 6000\, e^{-x^2}]\, dy = 6000\, y e^{-x^2}\Big|_{-x}^{x} = 12000\, x e^{-x^2}\,.$$

The outer integral can then be done by the substitution $u = e^{-x^2}$, leading to

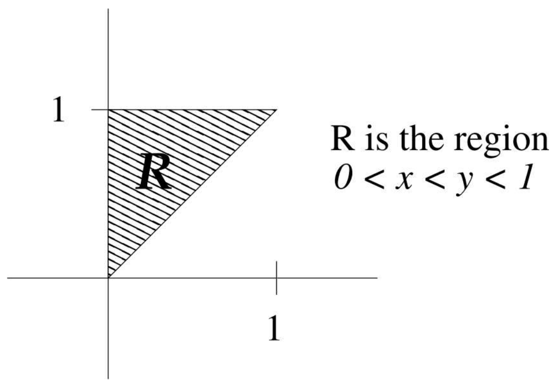
$$\int_0^\infty 12000\, x e^{-x^2}\, dx = -6000\, e^{-x^2}\Big|_0^\infty = 0 - (-6000) = 6000\,.$$

This is a good example of an integral which is not too hard one way but impossible the other. Try to integrate e^{-x^2} against dx rather than dy and you will be stuck at the first step! If you come across this, you will always want to switch the order of the integrals.

Averages

The average of a quantity over a region is just the total of the quantity divided by the size of the region.

EXAMPLE: What is the average of e^x over the triangular region where $0 \le x \le y \le 1$?



SOLUTION:

$$\int_0^1 \int_0^y e^x \, dx \, dy = \int_0^1 \left(e^x |_0^y \right) \, dy = \int_0^1 \left(e^y - 1 \right) \, dy = e - 2 \, .$$

The average value is the total, $e - 2$, divided by the area. The area is $1/2$ therefore the average value is $2(e - 2)$.

EXAMPLE: The cost of providing fiber optic service to a resident is proportional to the distance to the nearest hub, with constant of proportionality 5 dollars per meter. If a township is a square, two kilometers on a side, and there is a single hub in the center, what is the average of the service cost over this area?

SOLUTION: Make coordinates with the hub in the center. The township is the square $[-1000, 1000] \times [-1000, 1000]$, with x and y representing East-West displacement and North-South displacement in meters. The cost of providing service to the point (x, y) is given by $f(x, y) = 5\sqrt{x^2 + y^2}$. The average is therefore given by

$$\text{Ave.} = \frac{1}{1000^2} \int_R 5\sqrt{x^2 + y^2} \, dA = \frac{1}{1000^2} \int_{-1000}^{1000} \int_{-1000}^{1000} 5\sqrt{x^2 + y^2} \, dy \, dx \, .$$

If you can do this integral, you are smarter than I am. I tried it numerically with a 5×5 grid (I used midpoints and I used symmetry to restrict to the quadrant $[0, 1000] \times [0, 1000]$ in order to make my grid squares smaller) and got roughly \$3812 which is pretty close to what my computer tells me is the correct numeric value of \$3826.

Two-variable probability densities

It is often useful to consider a random pair of real numbers, that is, a random point in the plane. A probability density on the plane[4] is a nonnegative function $f(x, y)$ such that $\int f(x, y) dA = 1$. As before, the mean of the X variable is $\int x f(x, y)\, dA$ and the mean of the Y variable is $\int y f(x, y)\, dA$. Here are a couple of special cases.

Example: uniform density on a region. Let R be a finite region and let $f(x, y) = C$ on R and zero elsewhere. For this to be a probaiblity denstiy, the normalizing constant C must be the reciprocal of the area of R (that's because the integral of $1\, dA$ over R is just the area of R). For example, if R is the interior of the unit circle then C would be $1/\pi$. If R is the recangle $[a, b] \times [c, d]$ then $C = 1/((b - a)(d - c))$.

Example: planar standard normal distribution. Let $f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$. This has integral equal to 1 because it is the product of $(1/\sqrt{2\pi})e^{x^2/2}$ and $(1/\sqrt{2\pi})e^{y^2/2}$, which we already know integrate to 1 over the whole plane $(-\infty, \infty) \times (-\infty, \infty)$ because each one is just the one-variable standard normal density. This uses the "magic product formula".

A two-variable probability density corresponds to picking simultaneously two numbers X and Y such that the probability of finding the pair (X, Y) in some region A is equal to the integral of the density over the region A.

Example: A probability density on the rectangle $[0, 3] \times [0, 2]$ is given by Ce^{-x}. What is C, and what is the probability of finding the pair (X, Y) in the unit square $[0, 1] \times [0, 1]$?

[4] The integral, if it is over the whole plane, is technically an improper integral, but we won't worry about that; in all our examples either the density will be nonzero on just a finite region or it will be obvious that there is a limit as the region becomes infinite.

SOLUTION: We are integrating over a rectangle and Ce^{-x} is a product of $g(x) = Ce^{-x}$ and $h(y) = 1$. By the magic product formula, the integral is

$$\left(\int_0^3 Ce^{-x}\, dx \right) \times \left(\int_0^2 dy \right) = 2C \cdot \left(1 - e^{-3}\right) \,.$$

Therefore $C = \dfrac{1}{2(1-e^{-3})}$ which is just a shade over 1/2. Now using the product formula again to integrate over the unit square gives a probability of

$$\int_{[0,1]\times[0,1]} Ce^{-x}\, dA = C(1 - e^{-1}) = \frac{1 - e^{-1}}{2(1 - e^{-3})} \approx 0.3326.$$

11 Partial derivatives and multivariable chain rule

11.1 Basic defintions and the Increment Theorem

One thing I would like to point out is that you've been taking partial derivatives all your calculus-life. When you compute df/dt for $f(t) = Ce^{-kt}$, you get $-Cke^{-kt}$ because C and k are constants. The notation df/dt tells you that t is the variables and everything else you see is a constant. If we use the notation f' instead, then we are relying on your knowing which is the independent variable. It's usually called something like "t", not "C" or "k", but every now and then we end up computing df/dk or df/dC, so watch out! The only rule is: everyone should understand which is the independent variable.

So now, studying partial derivatives, the only difference is that the other variables aren't constants – they vary – but you treat them as constants anyway. It's not a big difference because really, what is a constant? It's always possible to imagine some quantity changing. Mathematically we just need to be precise about what is holding steady and what is changing. In this section, only one variable at a time will change. Then in the next section (chain rule), we'll change more than one independent variable at a time and keep track of the total effect on the independent variable.

We assigned plenty of MML problems on this section because the computations aren't much different than ones you are already very good at. You can read the basics in Section 14.3. I will include one example as a self-check; if you are not able to cover up the answer and figure it out pretty easily, then you need to go back and re-read Section 14.3.

EXAMPLE: Let $f(x, t, q) = \dfrac{e^q - 1}{1 + xtq}$. What is $\dfrac{\partial f}{\partial t}$ at the point $(3, 1, 1)$ and what does this quantity signify?

> Answer: treating everything other than t as a constant, by either the chain rule or the quotient rule you get $-xq(e^q - 1)/(1 + xtq)^2$. Evaluating at the point $(3, 1, 1)$ gives $-3(e - 1)/16$.
>
> This means that if t is changes by a small amount from 1 while x is held fixed at 3 and q at 1, the value of f would change by roughly $3(e - 1)/16$ times as much in the opposite direction.

The Increment Theorem

By now I'm sure you remember the linearization in one-variable. The value of $f(x)$ near the point $x = a$ is well approximated by $L(x) = f(a) + f'(a) \cdot (x - a)$. Suppose we now want to approximate $f(x, y)$ near a point (a, b) where we know the value. Suppose, in fact that we change only x but not y. Then we might as well treat y as a constant and write

$$f(x + \Delta x, y) = f(x, y) + (\Delta x) \cdot \frac{\partial f}{\partial x}(x, y) \,.$$

It's a partial derivative, not a total derivative, because there is another variable y which is being held fixed. Similarly, if we moved only y we would have

$$f(x, y + \Delta y) = f(x, y) + (\Delta y) \cdot \frac{\partial f}{\partial y}(x, y) \,.$$

I hope it doesn't seem like too much of a leap to say that if you move both x and y you'll get both of these effects:

$$f(x, y + \Delta y) = f(x, y) + (\Delta x) \cdot \frac{\partial f}{\partial x}(x, y) + (\Delta y) \cdot \frac{\partial f}{\partial y}(x, y) \,. \qquad (11.1)$$

Equation (11.1) is called the Increment Theorem in the textbook and appears as Theorem 3 on page 818 (Section 14.3). You might wonder whether it's OK to assume that you can just add the two effects from moving x and moving y. In fact, after you move x, you really should be computing the y increment according to the $\partial f / \partial y$ at the new location, $(x + \Delta x, y)$. However, it's only an approximntion anyway, and the new partial derivative is close enough to the old that the computation with the new partial derivative matches the computation with the old partial derivative to within the error you already introduce by linearizing.

Example: About how much does $x^2/(1 + y)$ change if (x, y) changes from $(10, 4)$ to $(11, 3)$? Here $\Delta x = 1$ and $\Delta y = -1$. We compute $f_x = 2x/(1 + y)$ and $f_y = -x^2/(1 + y)^2$ so so $f_x(10, 4) = 4$ and $f_y(10, 4) = -4$. Thus,

$$\Delta f \approx f_x \Delta x + f_y \Delta y = 4(1) + (-4)(-1) = 8 \,.$$

In fact, f changes from 20 to 30.25 so the 8 was kind of a crude estimate, but that's because Δx and Δy were pretty big. If we choose 0.1 and -0.1 instead, we get a linear estimate of $\Delta f = 0.8$ which is very close to the actual $0.818\ldots$.

Application: marginal rates

Suppose the cost of a proposed building is a function $f(A, q, \ell)$ where A is the area of usable space in square feet, q is an index of the quality (thickness of walls, gauge of wiring, level of insulation, quantity of lighting, etc.) and ℓ is a location parameter measuring, for example, the desirability of the location. The average cost per square foot for a given proposed building is, by definition, $f(A, q, \ell)/A$. However, this statistic is far less useful than the *marginal* cost per square foot, that is, $\partial f/\partial A$. That's because most decisions are about whether to put a few extra dollars into one of these categories or to trim a few bucks from another category. Therefore, it is most useful to know how many dollars more you will spend or save with each square foot, rather than what all the square footage costs that is already in all the proposals being compared.

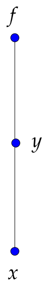
EXAMPLE: The total number P of people exposed to an recurring ad is a function of its market share, M, and the length of time, t, that stays in rotation[5]. The marginal increase in exposure per time run is $\partial f/\partial t$. The right time to yank the ad is when $v \cdot \partial f/\partial t$ drops below the cost per time to run the ad, where v is the value in dollars per unit of exposure. Note that the units match: v has units of dollars per exposure, $\partial f/\partial t$ has units of exposure per time and the cost to run the ad is priced in dollars per time: (\$/exp) (exp/t) = \$/t.

Note: the notion of marginal rates should already be familiar from univariate calculus. There isn't much added here, except to say that it makes sense to compute marginal rates when there are many quantities that could vary, by varying only one.

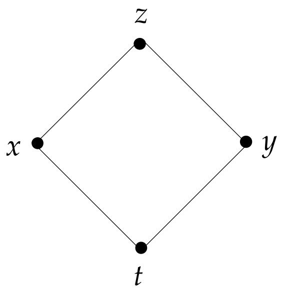
Branch diagrams

In applications, computing partial derivatives is often easier than knowing what partial derivatives to compute. With all these variables flying around, we need a way of writing down what depends on what. We do this by writing a branch diagram. Here are some common ones.

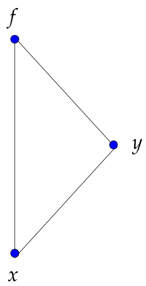
[5] It is not just the product of these because the longer it runs, the more redundancy there is in people seeing it multiple times.



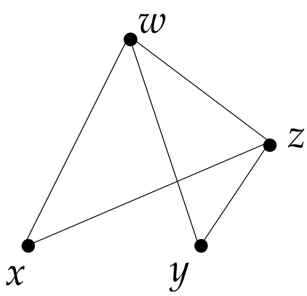
The branch diagram for the ordinary chain rule.



w is a function of x and y, both of which are functions of a single variable t (see page 823 of the textbook).



z depends on x and y but y is really a function of x



w is a function of x, y and z, but z is really a function of the other two.

Any variable at the top is an dependent variable. Any variable at the bottom is an independent variable; these drive the other variables and are the only ones we tweak directly. The variables in the middle are called *intermediate* variables. The independent variables drive them and they drive the dependent variables.

11.2 Chain rule

Think about the ordinary chain rule. A useful metaphor is that it is like a gear assembly[6]: y depends on u, which in turn depends on x. Each unit increase of x increases u by $u'(x)$ many units. Each unit increase of u inceases y by $y'(u)$ units. Therefore each unit increase in x produces $u'(x) \cdot y'(u)$ units increase in y. That's what's going on in the first branch diagram.

In the second diagram, there is a single independent indpendent variable t, which we think of as a gear driving both x and y, while both x and y drive z. I am going to try now to explain why

$$\frac{dy}{dt} = \frac{\partial y}{\partial u}\frac{du}{dt} + \frac{\partial y}{\partial v}\frac{dv}{dt}\,. \tag{11.2}$$

[6] OK, you got me, that's a simile not a metaphor.

When t increases by Δt, both u and v increase. The increases are roughly $(\Delta t)(du/dt)$ and $(\Delta t)(dv/dt)$ respectively. As we just saw at the end of the previous section (with the function $z(x, y)$) each increase in u produces an increase in y that is $\partial y/\partial u$ times as great. So the increase in u of $\Delta t \frac{du}{dt}$ gives gives an increase in y of roughly $\Delta t \frac{du}{dt}\frac{\partial y}{\partial u}$. Simultaneously, the increase in t has produced an increase in v which produces another increase in y of roughly $\Delta t \frac{dv}{dt}\frac{\partial y}{\partial v}$. Thus the total increase in y is roughly

$$\Delta t \left[\frac{\partial y}{\partial u}\frac{du}{dt} + \frac{\partial y}{\partial v}\frac{dv}{dt}\right] .$$

This means that the rate of change of y per change in t is given by equation (11.2). Note that we use partial derivative notation for derivatives of y with respect to u and v, as both u and v vary, but we use total derivative notation for derivatives of u and v with respect to t because each is a function of only the one variable; we also use total derivative notation dy/dt rather than $\partial y/\partial t$. Do you see why? Partial derivative notation would mean that t was changing while something else was being held fixed, which is not the case. Rather, all variables are functions of the single variable t.

That's the basic story. There are lots of variations, depending on how many independent variables there are (up till now there has been only one, all the others ultiimately being functions of the one), how many intermediate variables and how they are related.

Where to evaluate?

The one thing you need to be careful about is evaluating all derivatives in the right place. It's just like the ordinary chain rule. For example, in (11.2), the derivatives du/dt and dv/dt are evaluated at some time t_0. The partial derivative $\partial y/\partial u$ is evaluated at $u(t_0)$ and the partial derivative $\partial y/\partial v$ is evaluated at $v(t_0)$.

Example: Chain rule for $f(x, y)$ when y is a function of x

The heading says it all: we want to know how $f(x, y)$ changes when x and y change but there is really only one independent variable, say x, and y is a function of x. This is captured by the third of the four branch diagrams on the previous page. Applying the chain rule gives

$$\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \cdot y' \, . \tag{11.3}$$

The notation really makes a difference here. Both df/dx and $\partial f/\partial x$ appear in the equation and they are not the same thing!

Example: Derivative along an explicitly parametrized curve

One common application of the multivariate chain rule is when a point varies along a curve or surface and you need to figure the rate of change of some function of the moving point. The classical economics application is that price and quantity are moving together along the *demand curve* and we want to figure out how revenue changes along this curve (and in particular, we want to find where the revenue is maximized). In this section we solve the problem when the curve is known explicitly, saving the case of implicitly defined curves until we have discussed implicit differentiation.

Suppose a point varies along a curve as a function of time, and its coordinates are explicitly known: the coordinates at time t are $(x(t), y(t))$. The rate of change of the function $g(x, y)$ with respect to *time* along the curve is given by the formula we just computed: x and y are functions of t and g is a function of x and y, so

$$\frac{dg}{dt} = \frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial y}\frac{dy}{dt} \, . \tag{11.4}$$

I hope you realize this is the exact same equation as (11.2) but with the letter g in place of y, and x and y in place of u and v.

11.3 Implicit differentiation

The chain rule helps us to understand ordinary implicit differentiation. In Section 14.4 on page 826 the textbook re-explains finding the slope of an implicitly defined curve (first discussed in the textbook in Section 3.7). Here follows a quick recap of this.

Slope of an implicitly defined curve

Suppose a curve is defined by $F(x, y) = 0$. What is the slope of its tangent line? That's the same as asking, if we treat y as a function of x along the curve, what is dy/dx? This is just (11.3) run backwards – we know that $df/dx = 0$ and want to solve for y'. Differentiating the relation $F(x, y) = 0$ with respect to x, where y is an intermediate variable that is a function of x, the chain rule gives $0 = F_x + F_y\, dy/dx$. Solving for dy/dx gives (see page 826 of the textbook):

$$\frac{dy}{dx} = -\frac{F_x}{F_y}\,. \tag{11.5}$$

Derivative along an implicitly parametrized curve

Now suppose a curve is defined implicitly by $F(x, y) = 0$. How fast does the function $g(x, y)$ change along the curve? We had better decide: how fast does $g(x, y)$ change with respect to *what*? Suppose we treat y as a function of x along the curve and ask for dg/dx. Using the chain rule for this case (11.3)

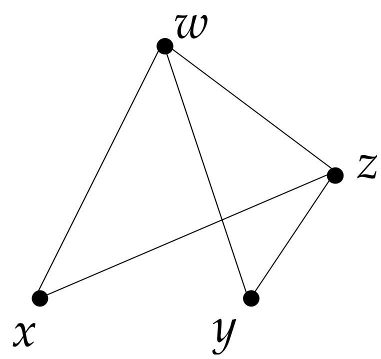
$$\begin{aligned}\frac{dg}{dx} &= \frac{\partial g}{\partial x} + \frac{\partial g}{\partial y}\frac{dy}{dx} \\ &= \frac{\partial g}{\partial x} - \frac{\partial g}{\partial y}\frac{\partial F/\partial x}{\partial F/\partial y}\,.\end{aligned}$$

In the last line, we used the expression for dy/dx given by implicit differentation (11.5).

Implicitly defined surfaces

This is just like curves defined by an equation, only now there are three variables. Any equation $F(x, y, z) = 0$ defines a surface. If any two vary freely, the third changes as a function of the other two. When this happens, we can ask for the rate of change of one with respect to another. What should $\partial z/\partial x$ mean in this context? It means: consider z as a function of x and y, then find out the rate of change in z when x varies, y is held constant, and z changes in order still to satisfy the equation. Please take a monent to think this through now.

Computationally, how do we find $\partial z/\partial x$ when $F(x, y, z) = 0$? We differentiate, keeping in mind the branch diagram. Letting w denote $F(x, y, z)$, it is the same as one we have seen before:



The variables vary in such a way that w remains at zero. Taking the partial derivative with respect to x of the equation $w = 0$ gives

$$0 = \frac{\partial w}{\partial x}(x, y, z) = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial z} + \frac{\partial z}{\partial x}\,.$$

Solving for $\partial z/\partial x$ we see that

$$\frac{\partial z}{\partial x} = \frac{-F_x}{F_z}\,.$$

This looks exactly the same as for two variables, x and z only; compare to equation (11.5). This is not a coincidence. If z is a function of x and y and we hold y constant, then y is playing a similar role to the constant k in the function e^{kx}. The problem really does reduce to the two variable problem. Let's try it on Example 4 from Section 14.3 of the textbook.

EXAMPLE: Find $\partial z/\partial x$ when the equation $F(x, y, z) = x + y + \ln z - yz = 0$ defines z as a function of x and y. We compute $F_x = 1$ and $F_z = 1/z - y$ therefore

$$\frac{\partial z}{\partial x} = \frac{-1}{1/z - y} = \frac{z}{yz - 1}\,.$$

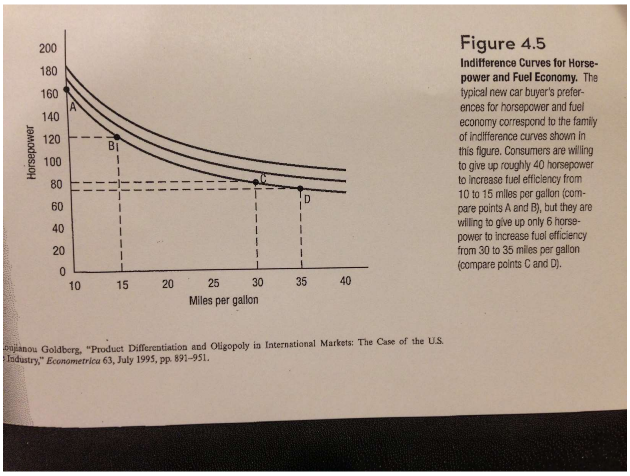
You should compare this to how the book does it (page 813); I think this way is simpler than the book's but either is OK.

11.4 Featured application: indifference curves

Remember level curves from our first day of multivariate calculus? They're back, in an economic application, under the name of "indifference curves". Suppose that the independent variables x and y represent quantities of two different things that will rival each other for importance in a single scenario.

Example 1: x is the horsepower of a car and y is its MPG.

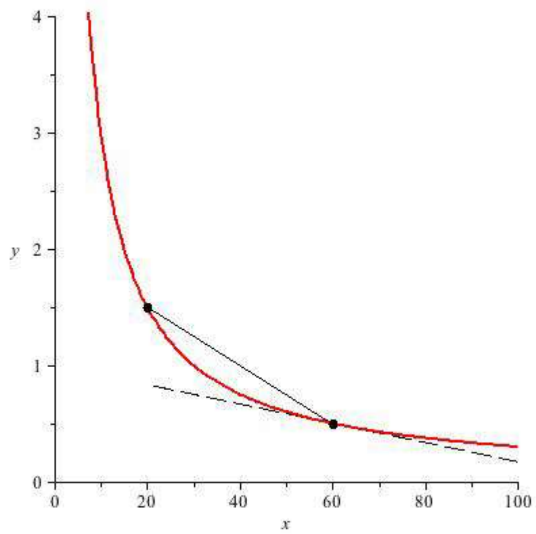
Example 2: x is ounces of pizza at a meal and y is pints of FroYo.



An indifference curve is a set of points in the x-y plane corresponding to bundles that the agent (often a consumer) likes equally well. The two examples above are taken from Berheim and Whinston (current textbook for BEPP 250). The indifference curve for horsepower versus fuel economy is taken from actual data. The indifference curve for pizza versus FroYo is a made up model. In either case, the indifference curves are just level contours for a utility function $u(x, y)$. The food example uses the utility function $u(x, y) = xy$ and shows indifference curves of $xy = 10, xy = 20$ and $xy = 30$.

Indifference curves are important for several reasons, one of which is that they describe incentives and reactions to changes in the quantities x and y. The *marginal rate of substitution* is the amount of x an agent would be willing to give up in order to increase y by one unit. This is not a static quantity, rather it depends on the present levels of x and y. If a group of diners has 10 pints of FroYo and only three ounces of pizza, they will not be willing to give up much pizza for one more pint of FroYo, whereaas a group with 60 ounces of pizza and half a pint of FroYo might well give up a lot of pizza for a pint of FroYo.

Two points on the same indifference curve, such as $(60, 1/2)$ and $(20, 3/2)$, determine an equivalence of utility. The slope of the line segment between these two points is a ratio for a trade the agent is willing to make in either direction (see the straight line in the figure). But the point $(20, 3/2)$ is quite far from $(60, 1/2)$ and does not represent the rate of substitution if the consumer is able to make continuous small adjustments. As the point (x, y) on the curve $u(x, y) = 30$ approaches $(60, 1/2)$, the slope of the line segment approaches the slope of the tangent line to the curve $u(x, y) = 30$ at $(60, 1/2)$ (dashed line in the figure).

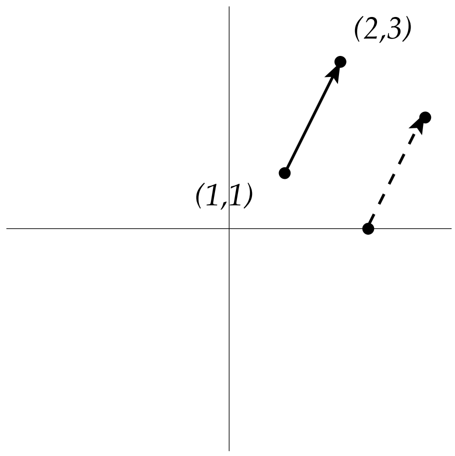


Mathematically, the marginal rate of substitution is defined to be the negative of the slope of this tangent line (negative because the slope represents one quantity going down while the other goes up). This slope is just dy/dx, which we know how to compute via implicit differentiation. In the pizza and FroYo example, the level curve is $xy = 30$ and implicit differentiation gives $y + x(dy/dx) = 0$. Thus $dy/dx = -y/x$. At the point $(60, 1/2)$, this gives a marginal rate of substitution of $1/120$ pint of FroYo per ounce of pizza. On the other hand, at the point $(3, 10)$, the marginal rate of substitution is $10/3$ pints of FroYo per ounce of pizza. Whether or not you think xy is a reasonable utility function for this scenario, this model sheds light on consumer behavior and how to model it.

12 Gradients and optimization

12.1 Vectors

Think of a vector as an arrow drawn from one point in the plane or three dimensional space to another. The arrow from $(1, 1)$ to $(2, 3)$ is shown in the figure. The only tricky thing about the definition is that we don't care where the arrow is drawn, we only care about its magnitude (length) and direction. So for example the dashed arrow represents the same vector, started at the point $(5/2, 0)$ instead of $(1, 1)$. In other words, the vector represents the *move* from the beginning to the end of the arrow, regardless of the absolute location of the beginning point.

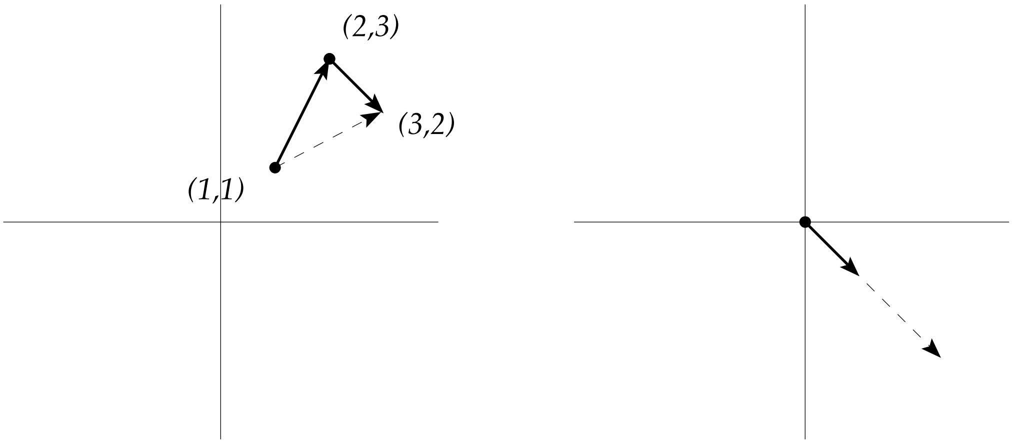


The vector of unit length in the x-direction is called $\hat{\mathbf{i}}$, the vector of unit length in the y-direction is called $\hat{\mathbf{j}}$, and, if we're in three dimensions, the vector of unit length in the z-direction is called $\hat{\mathbf{k}}$. A vector that goes a units in the x-direction and b units in the y-direction is denoted $a\hat{\mathbf{i}} + b\hat{\mathbf{j}}$. It's called that because you can add vectors and multiply them by real numbers (see definition below). For example, the vector in the picture should be written $\hat{\mathbf{i}} + 2\hat{\mathbf{j}}$.

> **Definition of adding vectors.** First make one move, then make the other. You can do this by sliding one of the arrows (don't rotate it!) so it starts where the other one ends, then following them both. If you add $a\hat{\mathbf{i}} + b\hat{\mathbf{j}}$ to $c\hat{\mathbf{i}} + d\hat{\mathbf{j}}$ you get $(a + c)\hat{\mathbf{i}} + (b + d)\hat{\mathbf{j}}$.

Definition of multiplying a vector by a real number. Don't change the direction, just multiply the length. As a formula: multiply $a\hat{\mathbf{i}} + b\hat{\mathbf{j}}$ by c you get $ac\hat{\mathbf{i}} + bc\hat{\mathbf{j}}$. This easy formula hides an important fact: if you mutliply both the $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ coefficients by the same real number, the direction doesn't change. That's why the two vectors in the right-hand figure below are on top of each other.

The left-hand side of the figure below shows the vector $\hat{\mathbf{i}} + 2\hat{\mathbf{j}}$ being added, tip to tail, to the vector $\hat{\mathbf{i}} - \hat{\mathbf{j}}$. The result is the vector $2\hat{\mathbf{i}} + \hat{\mathbf{j}}$ show by the dotted arrow. In the right-hand figure, the vector $\hat{\mathbf{i}} - \hat{\mathbf{j}}$ is multiplied by the real number $\sqrt{6}$ which is a little under 21/2.



The length of a vector can be computed by the Pythagorean Theorem. The length of $a\hat{\mathbf{i}} + b\hat{\mathbf{j}}$ is $\sqrt{a^2 + b^2}$. For example, the vector $\hat{\mathbf{i}} + 2\hat{\mathbf{j}}$ which appears in the previous figures has length $\sqrt{5}$. The length of the vector $\mathbf{v}$ is denoted $|\mathbf{v}|$. A *unit vector* is any vector whose length is 1. Often we want to know a unit vector in a given direction: what vector, having the same direction as $\mathbf{v}$, has length 1? Answer: divide $\mathbf{v}$ by $|\mathbf{v}|$ (that is, multiply $\mathbf{v}$ by the reciprocal of its length). Self-check: what is the unit vector in the direction of our favorite example vector, $\mathbf{v} = \hat{\mathbf{i}} + 2\hat{\mathbf{j}}$? The answer is posted in a link on Canvas (first student who actually wants to look at it, tell me and I'll activate the link).

The dot product

The *dot product* of the vectors $a\hat{\mathbf{i}} + b\hat{\mathbf{j}}$ and $c\hat{\mathbf{i}} + d\hat{\mathbf{j}}$ is defined to be the number (it's not a vector!) $ac + bd$. You'll see next class why this quantity is important. The last thing you need to know is a fact: the dot product of two vectors $\mathbf{v}$ and $\mathbf{w}$ is equal to the product of the lengths times the cosine of the angle $\alpha(\mathbf{v}, \mathbf{w})$ between them:

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}||\mathbf{w}| \cos \alpha(\mathbf{v}, \mathbf{w}) \tag{12.1}$$

Again, there is something important hidden in the content of this formula. You already know one way of computing the dot product: multiply corresponding components and add them. The formula gives you another way. The first way is algebraic. The second way is completely geometric: you could do it by seeing only the picture. The dot product theorem says that these two computations produce the same result. Take a minute to register this, because it will come up in applicaitons, problem sets and, yes, exams.

Parallel vectors

Vectors in the same direction are called *parallel.* How do you tell whether $\mathbf{v} = a\hat{\mathbf{i}} + b\hat{\mathbf{j}}$ is parallel to $\mathbf{w} = c\hat{\mathbf{i}} + d\hat{\mathbf{j}}$? This is the same as saying you can multiply one vector by real number to get the other. This is the same as asking when the fraction c/a is equal to d/b. To test this you crossmultiply, arriving at the condition

$$ad - bc = 0\,. \tag{12.2}$$

Three or more dimensions (optional paragraph)

In three dimensions a generic vector will be the sum of three components: $a\hat{\mathbf{i}}+b\hat{\mathbf{j}}+c\hat{\mathbf{k}}$. The basic definitions are still the same. A vector $\mathbf{v}$ is still defined as having a length and a direction. Both the algebraic and the geometric formulae for the dot product look analogous to they way they looked in two dimensions and give the same answer. Addition of vectors and multiplication of a vector by a real number still have both an algebraic and a geometric definition that give the same result. In fact, you define vectors in any dimension. You can't visualize it, and you run out of letters after $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$, but all the math still works the way it did in two dimensions. Our treatment is very minimal: we will stick to two dimensions. If vector calculus intrigues you then consider taking Math 114.

12.2 The gradient

Let z be a function of x and y. Think of this for now as the elevation at a point x units east and y units north of a central point. Pick a point (x_0, y_0), let $a = (\partial z/\partial x)(x_0, y_0)$ and let $b = \partial z/\partial y)(x_0, y_0)$. Using these we can figure out the rate of elevation increase for a hiker traveling on the path $(x(t), y(t))$. By the multivariate chain rule, if the hiker is at the position (x_0, y_0) at some time t_0, then the rate of increase of the hiker's elevation at time t_0 will be $ax'(t) + by'(t)$ evaluated at $t = t_0$.

Here's the important point. If we calculate a and b just once, we can figure out the rate of elevation gain of any hiker traveling with any speed in the x- and y-directions. The vector $a\hat{\mathbf{i}} + g\hat{\mathbf{j}}$ is called the *gradient* of z at the point (x_0, y_0) and is denoted $\nabla z(x_0, y_0)$ or just ∇z. This definition is given in a box in the middle of page 833 in Section 14.5 of the textbook:

$$\nabla z(x_0, y_0) = \frac{\partial z}{\partial x}(x_0, y_0)\,\hat{\mathbf{i}} + \frac{\partial z}{\partial y}(x_0, y_0)\,\hat{\mathbf{j}}\,.$$

This leads to the idea of the *directional derivative*: what is the rate of elevation gain per unit traveled in any direction? The key here is "per unit traveled". The unit vector $\mathbf{w}$ in the direction making an angle of θ with the positive x-direction is $(\cos\theta)\hat{\mathbf{i}} + (\sin\theta)\hat{\mathbf{j}}$. Therefore, a hiker traveling at unit speed in this direction gains elevation at the rate of $a\cos\theta + b\sin\theta$. That's the dot product $\nabla z \cdot \mathbf{w}$. THIS IS THE MAIN REASON WE COVER VECTORS AND DOT PRODUCTS IN THIS COURSE.

Here are some conclusions you can draw from all of this. Let $L = |\nabla z(x_0, y_0)|$ be the length of the gradient vector of z at the point (x_0, y_0). Now consider all directions the hiker could possibly be traveling: which one maximizes the rate of elevation gain? Let α be the angel between the gradient vector and the hiker's direction in the x-y plane. We have just seen that the rate of elevation gain per unit motion in the direction $\mathbf{w}$ is $\nabla z \cdot \mathbf{w}$. The length of ∇z is L and the length of $\mathbf{w}$ is 1, so by formula (12.1), the dot product is $L\cos\alpha$. This cosine is at most 1 and is maximized when the angle is zero, in other words, when the hiker's direction is parallel to the gradient vector. In that case the directional derivative is L. If the hiker is going in a direction making an angle α with the gradient then the rate of elevation gain per unit distance traveled is $L\cos\alpha$. If α is a right angle then this rate is zero. We can summarize these observations in a theorem, which constitutes more or less the "Properties of the directional derivative" stated in a box on page 834.

Gradient Theorem:

> (i) The direction of greatest increase of a function $z(x, y)$ at a point (x_0, y_0) is the direction of its gradient vector $\nabla z(x_0, y_0)$. The rate of increase per unit distance traveled in that direction is the length of the gradient vector which is given by
>
> $$L = \sqrt{\frac{\partial z}{\partial x}(x_0, y_0)^2 + \frac{\partial z}{\partial y}(x_0, y_0)^2}\,.$$
>
> (ii) In general the directional derivative in a direction making angle α with respect to the gradient direction is equalt to $L \cos \alpha$.
>
> (iii) In particular, when α is a right angle, we see that the rate of elevation increase in direction α is zero.

This theorem is, more or less what's in the box on page 834 entitled "Properties of the directional derivative". Mull it over for a minute. By computing partial derivatives, we can stake out the direction of maximum ascent, and it will have the property that the direction of zero elevation gain is at right angles to it (also, the direction of maximal descent is exactly opposite). Remember level curves? Along these, the elevation is constant. Therefore, traveling in these directions makes the rate of elevation gain zero. We see that the tangent to the level curve must be in the zero gain direction, that is, perpendicular to the gradient. This is shown in Figure 14.31 on page 835. A real life illustration is shown in the picture on page 831 of the textbook. A contour map shows contours of an actual mountainside in Yosemite National Park. These are perpendicular to the directions of steepest ascent and descent. You can see this because streams typically flow in the directions of steepest descent. The streams and the level contours are marked on the map and do, indeed, look perpendicular.

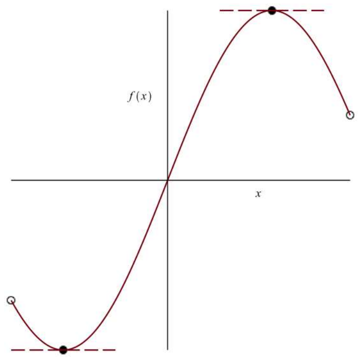
Some rules for computing

We won't need a lot of rules for computing gradients because we'll always be able to compute them by hand but it is good to look them over once. They're collected in a box on page 836 of the textbook. Basically all the rules that work for derivatives work for gradients because in each component separately (the $\hat{\mathbf{i}}$ component, etc.) the gradient is a kind of all-encompassing partial derivative, and partial derivatives obey these laws.

12.3 Optimization

One-paragraph review of univariate optimization

Here is a brief review of optimization in one variable (also known as solving max-min problems) which you can skip if you don't need. Suppose you want to find the maximum of a differentiable function f on an interval $[a, b]$. If the maximum occurs at a point x in the interior of the interval then $f'(x)$ must be zero (if $f'(x) > 0$ then a point just to the right of x will have a higher value of f, whereas if $f'(x) < 0$ then a point just to the left of x will have a higher value of f). Therefore, to find the maximum, you list all the critical points (points where $f' = 0$ and both the endpoints, and see where among these points f has the greatest value. It is the same with minima. List all the critical points of f and the endpoints, and determine the least value of f among this list of points.



The maximum and minimum of f on the interval shown must occur either at the critical points (solid dots) or at the endpoints (open dots). In fact the maximum value (upper dashed line) and minimum value (lower dashed line) occur at critical points in this case.

Optimization along a curve

Now switch gears and consider a function $f(x, y)$ of two variables. There are two kinds of optimization problems that commonly occur. One is to find the maximum or minimum of f on a curve. The second is to find the maximum or minimum of f over a region in the plane.

Conceptually, optimization along a curve is easy: read f "as you go along the curve";

find the critical points where the derivative of the readout is zero; the maximum will have to occur at one of these places; check them all. Computationally, the tricky part is to describe the curve in equations, then use those equations to compute the derivative along the curve.

The description of a curve γ can take one of three forms. It could be given by some function $y = g(x)$. It could be given parametrically by $((x(t), y(t))$. Finally, and most commonly, γ could be given implicitly, meaning it is the solution set to the equation $H(x, y) = 0$ for some function H. We treat these in the order: parametric, function, implicit, because each computation relies on the previous one.

Parametric case: the derivative along $(x(t), (y(t))$.

If the curve γ is paramterized as $(x(t), y(t))$, then the derivative of f along γ is just $\nabla f \cdot \mathbf{v}$ where $\mathbf{v}$ is the velocity vector $x'(t)\hat{\mathbf{i}} + y'(t)\hat{\mathbf{j}}$. In this case, finding the points where the derivative of f along γ vanishes boils down to solving

$$x'(t)\frac{\partial f}{\partial x} + y'(t)\frac{\partial f}{\partial y} = 0\,. \tag{12.3}$$

Self-check: what does it mean that the derivative of f along the curve $(x(t), y(t))$ is given by (12.3)? This formula computes the rate of change of what with respect to what?

Function case: the derivative along $y = g(x)$.

If γ is paramtrized by $y = g(x)$ then you can use the parametric description $x = x, y = g(x)$ so that this equation becomes

$$\frac{\partial f}{\partial x} + g'(x)\frac{\partial f}{\partial y} = 0\,. \tag{12.4}$$

Self-check: again, this is the rate of change of what with respect to what?

Implicit case: the derivative along $H(x, y) = 0$.

Finally, suppose that γ is given implicitly by $H(x, y) = 0$. Recall that we know how to find the slope dy/dx of the tangent line to the level curve $H(x, y) = 0$. By implicit differentiation, we computed $dy/dx = -H_x/H_y$. Therefore we can apply equation (12.4) with $g'(t) = -H_x/H_y$. We get $\partial f/\partial x - (H_x/H_y)\partial f/\partial y = 0$, which simplifies slightly to

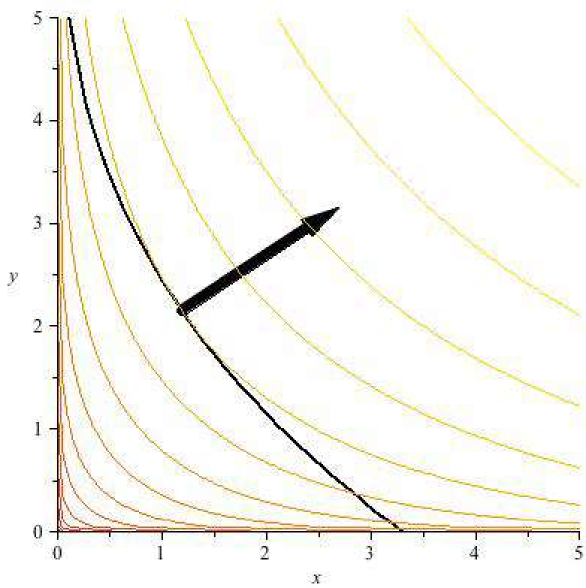
$$H_y\frac{\partial f}{\partial x} - H_x\frac{\partial f}{\partial y} = 0\,. \tag{12.5}$$

IMPORTANT GEOMETRIC INTERPRETATION OF (12.5):

The gradient of H is $H_x\hat{\mathbf{i}} + H_y\hat{\mathbf{j}}$. The gradient of f is $f_x\hat{\mathbf{i}} + f_y\hat{\mathbf{j}}$. The test for these to be parallel is given by applying (12.2) to these two vectors. This results precisely in (12.5). In other words:

> The critical points of f along a level curve of H are those points where the gradients of f and H are parallel.

Pictorial example: The figure shows a black constraint curve, $H(x, y) = 0$, along with contours for another function $f(x, y)$. The maximum of f along the curve $H(x, y) = 0$ is the place where the level curves, when you move from higher to lower, just hit the black curve. At this point, the curves are tangent and the gradients are parallel. The single arrow represents the directions of both gradients.



12.4 Optimization over a region

If the maximum of f on the region R occurs at an interior point, then both partial derivatives must vanish there. Why? If one of the partials, say $\partial f/\partial x$ is positive, then the value of the function is just to the right is greater. If, say $\partial f/\partial y$ is negative, then the value of f just below is greater. And so forth. Asking that both partial derivatives vanish is the same as asking that the gradient vanishes. Therefore, we have the following procedure.

> To find the maximum, find all the places inside R where the gradient vanishes, compute f at these places and take the maximum among these, and compare to the maximum of f on the boundary of R.

Note: the last part, finding the maximum on the boundary of f, unless this boundary is very simple, relies on knowing how to do constrained optimization.

Example: What is the maximum of the function $x + 2y$ on the region R where the unit circle intersects the first quadrant?

Solution: The gradient of $f(x) = x + 2y$ is the vector $\hat{\mathbf{i}} + 2\hat{\mathbf{j}}$. This never vanishes so the maximum of f is NOT in the interior of the unit circle. The boundary of the region R is made of three pieces: the line segment on the x-axis from the origin to $(1, 0)$; the line segment on the y-axis from the origin to $(0, 1)$; the arc of the unit circle in the first quadrant. The maximum of f on each line segment occurs at the endpoint away from the origin, with values 1 and 2 respectively.

To find the maximum of $x + 2y$ on the arc $x^2 + y^2 = 1$, we compute the gradient of $x^2 + y^2$ which is $2x\hat{\mathbf{i}} + 2y\hat{\mathbf{j}}$. This is parallel to $\nabla f = \hat{\mathbf{i}} + 2\hat{\mathbf{j}}$ when the cross-multiple $2(2x) - 1(2y) = 0$. This happens when $y = 2x$. Plugging in $x^2 + (2x)^2 = 1$, we find that $x = 1/\sqrt{5}$ and $y = 2/\sqrt{5}$. There, $f(x, y) = x + 2y = 1/\sqrt{5} + 4/\sqrt{5} = 5/\sqrt{5} = \sqrt{5} \approx 2.23606$. This beats all the other maxima, therefore the global maximum of $x + 2y$ in the unit circle in the first quadrant is $\sqrt{5}$ and is achieved at $(1/\sqrt{5}, 2/\sqrt{5})$.

Example: Where is the maximum of $f(x, y) = x/(1+x^2+y^2)$ on the disk of radius 2? The critical points on the interior are where both partial derivatives of f vanish. The partial derivatives, when expressed with denominator $(1 + x^2 + y^2)^2$ have respective numerators $-2xy$ and $1+y^2-x^2$. Setting these equal to zero gives the points $(\pm 1, 0)$; here $f(x, y) = \pm 1/2$. On the boundary, the denominator is 5 so f can be no more thna $2/5$, therefore the overall maximum is $1/2$ at the point $(1, 0)$.

Application

Let's go back to the pizza and FroYo example from Unit 11.4, but without numbers. Let $H(x, y)$ be the utility of a consumer who gets x ounces of pizza and y pints of FroYo. Let $f(x, y)$ be the cost to me of producing x ounces of pizza and y pints of Froyo. For my ten dollar family bargain, I need to offer a pair that is on the curve $H(x, y) = c$ because that's what Burger Chef is offering and I will lose customers if my pizza-FroYo combo is less desirable than theirs. But my function f is different from Burger Chef's because my prodution line is different. Question: what bundle should I offer?

In mathematical terms, What value of (x, y) on the curve $H(x, y) = c$ minimizes $f(x, y)$? We just saw the answer to that: it is either an endpoint of the curve or a place where ∇f is parallel to ∇H. Let's interpret the parallel gradients in economic terms. Parallel gradients at a point occur when the tangent lines to the level curves are the same at that point. These tangents tell me the marginal rate of substitution. Remember the FroYo example. The tangent to $H(x, y) = 30$ at the point $(60, 1/2)$ tells me the marginal rate of substitution. Consumers at this point are indifferent between another ounce of pizza and another $1/120$ point of FroYo. The tangent to the level curve of f at this point tells me the rate of substitution for costs: how many extra pints of FroYo can I make from the cost savings on each fewer ounce of pizza? If the two slopes are not the same, then I can slide along the customers indifference curve one direction or the other, decreasing my costs while maintaining the same customers. The only way I can be at the minimim cost point on the consumers' indifference curve is to be at a point where the slopes are parallel.

EXAMPLE: Using the numbers $H(x, y) = xy$ from the original pizza and FroYo example, suppose my cost function is a simple linear function: it coses 10 cents to produce each ounce of pizza and \$1 for each pint of FroYo. Thus $f(x, y) = (0.1)x + y$. The gradient of a linear function is constant: $\nabla f = (1/10)\hat{\mathbf{i}} + \hat{\mathbf{j}}$. The gradient of H is $y\hat{\mathbf{i}} + x\hat{\mathbf{j}}$. These are parallel when $y - x/10 = 0$. At what point on the curve $H(x, y) = 30$ does this occur? We solve

$$\begin{aligned} x &= 10y \\ xy &= 30 \end{aligned}$$

to get $y = \sqrt{3}$ and $x = 10\sqrt{3}$. Look up the approximate value $\sqrt{3} = 1.732\ldots$ on your cheatsheet. In other words, the optimum combo meal for me to sell is (roughly) a 17 and a third ounce pizza and a pint and three quarters of FroYo.