Notes for Calculus Lifesaver

CHAPTER 1 Functions, Graphs, and Lines

1.1. Functions

● polynomials; exponentials; logarithms; trigonometric; Intermediate Value Theorem; Max-Min Theorem; Rolle's Theorem; The Mean Value Theorem; Mean Value Theorem for integrals; sphere V; quadratics; cone v; velocity; Euler's identity; derivatives involving Trig functions; overview of Limits; products-to-sums identities; geometric series; the Taylor series; five Maclaurin series;

$$\sum_{j=a}^{b}\left(f(j)-f(j-1)\right)=f(b)-f(a-1)$$

$$\sum_{j=1}^{n}1=n$$

$$\text{sum of first } n \text{ natural numbers (telescoping series): } \sum_{i=1}^{n} i=\frac{n(n+1)}{2}$$

$$\sum_{j=1}^{n}(2j-1)=n^2$$

$$\text{sum of the squares of first } n \text{ natural numbers: } \sum_{i=1}^{n} i^2=1^1+2^2+\cdots+n^2=\frac{n(n+1)(2n+1)}{6}$$

$$\sum_{i=0}^{n}(2i+1)^2=1^1+3^2+\cdots+(2n+1)^2=\sum_{i=0}^{n}(4i^2+1+4i)=\frac{(n+1)(2n+1)(2n+3)}{3}$$

$$\boxed{\begin{gathered} a^3-b^3=(a-b)(a^2+ab+b^2) \\ a^3+b^3=(a+b)(a^2-ab+b^2) \\ a^2-b^2=(a-b)(a+b) \end{gathered}}$$

Definite integrals:

$$\int_a^b f(x)\,dx=\lim_{\text{mesh}\to 0}\sum_{j=1}^{n} f(c_j)(x_j-x_{j-1})$$

where $a=x_0<x_1<\cdots<x_{n-1}<x_n=b$ and c_j is in $[x_{j-1},x_j]$ for each $j=1,\cdots,n$.

Indefinite integrals:

$$\int f(x)dx=F(x)+C$$

Improper integral; Finding Taylor series.

● **R**: the set of all real numbers: The real numbers include all the rational numbers **[Q]**, such as the integer **[Z]**[N] -5 and the fraction $4/3$, and all the irrational numbers, such as $\sqrt{2}$ (1.41421356…, the square root of 2, an irrational algebraic number). Included within the irrationals are the transcendental numbers, such as π (3.14159265…). See **Figure 1**.

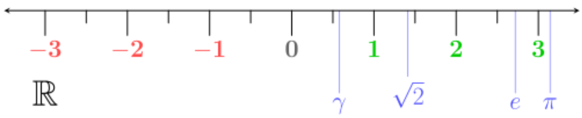


Figure 1. **[R]**: Real number; **[Q]**: rational numbers; **[Z]**: integers; **[N]**: natural numbers

1.1.1. Interval notation

● $\{x: 2 \leq x < 5\} = [2,5)$

1.1.2. Finding the domain (like x)

● The denominator of a fraction can't be zero. $=!\ 0$

● You can't take the square root (or fourth root, sixth root, and so on) of a negative number. ≥ 0

● You can't take the logarithm of a negative number or of 0. > 0

● $(-8, 13] \setminus \{2\}$: the domain is the set $(-8;\ 13]$ except for the number 2

1.1.3. Finding the range using the graph

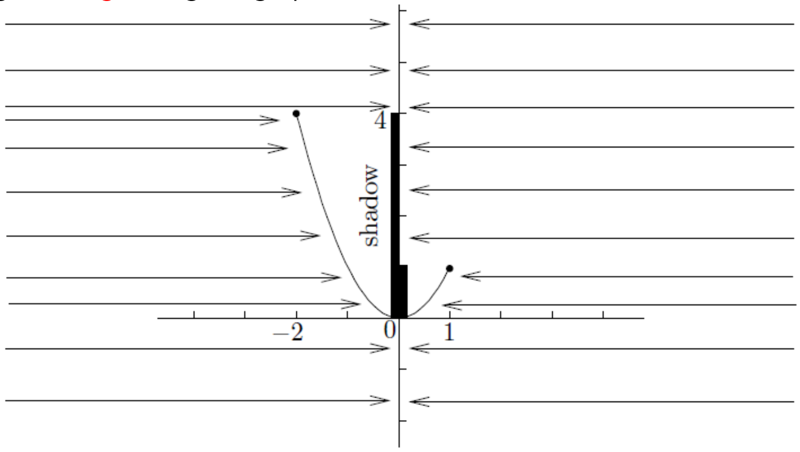


Figure 2. the range is $[0,4]$

● Remember, the codomain of any function we look at will always be the set of all real numbers.

● codomain $\neq$ range.

1.1.4. The vertical line test

● How to judge whether it is a function? -> check the vertical lines: whether two or more points on the graph can lie on the same vertical line, see **Figure 3 and 4**.

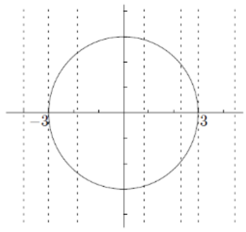


Figure 3. Not a function ($x^2 + y^2 = 9$)

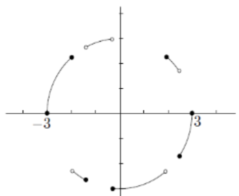


Figure 4. A function

1.2. Inverse Functions

● Start with a function f such that for any y in the range of f, there is exactly one number x such that $f(x) = y$. That is, different inputs give different outputs. Now we will define the inverse function f^{-1}.

● The domain of f^{-1} is the same as the range of f.

● The range of f^{-1} is the same as the domain of f.

● The value of $f^{-1}(y)$ is the number x such that $f(x) = y$. So,

if $f(x) = y$; then $f^{-1}(y) = x$.

1.2.1. The horizontal line test

● How to judge whether the function has an inverse function? -> check the horizontal lines: whether even one horizontal line intersects the graph more than once, see **Figure 5 and 6**.

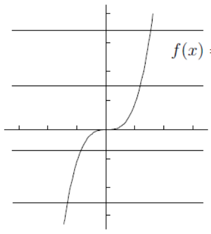


Figure 5. Inverse function existed

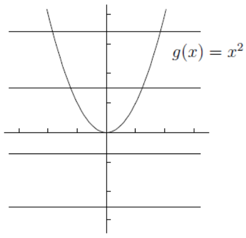


Figure 6. Inverse function doesn't exist

1.2.2. Finding the inverse

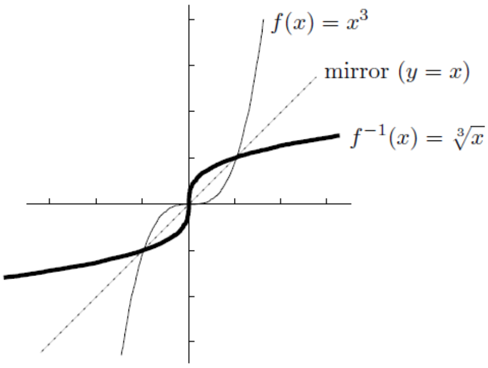


Figure 7. Draw the line $y = x$ to find the inverse

1.2.4. Inverses of inverse functions

If the domain of a function f can be restricted so that f has an inverse f^{-1}, then

● $f(f^{-1}(y)) = y$ for all y in the range of f; but

● $f^{-1}(f(x))$ may not equal x; in fact, $f^{-1}(f(x)) = x$ only when x is in the restricted domain.

1.3. Composition of Functions

● $f(x) = h(g(x))$ can be expressed as $f = hOg$, so

$$f(x) = m(k(j(h(g(x))))) \text{ can write } f = mOkOjOhOg$$

1.4. Odd and Even Functions

● Even Functions: $f(-x) = f(x)$

● Odd Functions: $f(-x) = -f(x)$

● Remember, odd functions must pass through the origin if they are defined at 0

● The product of two odd functions is always an even function, the product of two even functions is always even, and also that the product of an odd and an even function must be odd.

1.5. Graphs of **Linear** Functions

● $f(x) = mx + b$, the slope is m, and the y-intercept is b. See **Figure 8**.

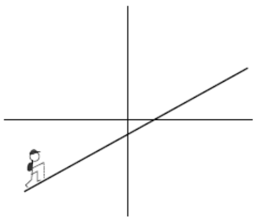


Figure 8. $f(x) = mx + b$

● the *point-slope* form of a linear function

> If a line goes through (x_0, y_0) and has slope m,
> then its equation is $y - y_0 = m(x - x_0)$.

1.6. Common Functions and Graphs

● **Polynomials:** these are functions built out of nonnegative integer powers of x. You start with the building blocks $1, x, x^2, x^3$, and so on, and you are allowed to multiply these basic functions by numbers and add a finite number of them together.

● The highest number n such that x^n has a nonzero coefficient is called the *degree* of the polynomial.

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x^1 + a_0$$

where a_n is the coefficient of x^n, a_{n-1} is the coefficient of x_{n-1}, and so on down to a_0, which is the coefficient of 1.

● **Quadratics:** $p(x) = a_2 x^2 + a_1 x^1 + a_0$, it's easier to write the coefficients as a, b, and c, so we have $p(x) = ax^2 + bx + c$. Quadratics has two, one, or zero (real) roots, depending on the sign of the *discriminant* Δ. $\Delta = b^2 - 4ac$, There are three possibilities. If $\Delta > 0$, then there are two roots; if $\Delta = 0$, there is one root (called a *double root*); and if $\Delta < 0$, then there are no roots. In the first two cases, the roots are given by

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

● An important technique for dealing with quadratics is *completing the square*, like

$$2x^2 - 3x + 10 = 2\left(x^2 - \frac{3}{2}x + 5\right) = 2\left(\left(x^2 - \frac{3}{2}x + \frac{9}{16}\right) + 5 - \frac{9}{16}\right) = 2\left(\left(x - \frac{3}{4}\right)^2 + \frac{71}{16}\right)$$

● **Rational functions:** these are functions of the form

$$\frac{p(x)}{q(x)}$$

where p and q are polynomials. Some simplest examples are **Figure 9**:

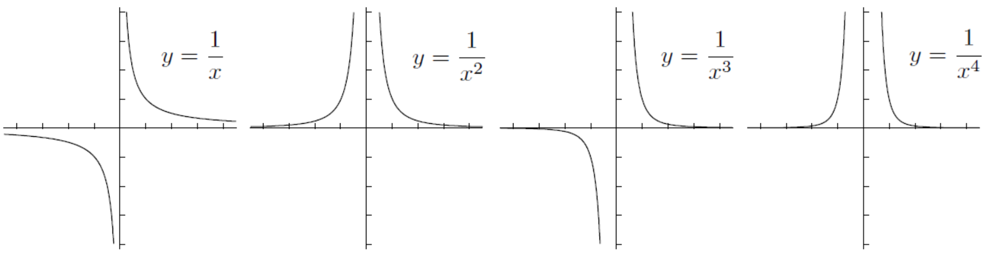


Figure 9. Some simplest rational functions

● **Exponentials and logarithms:** exponentials are $y = b^x$, the domain is the whole real line, the y-*intercept* is 1, the range is $(0, \infty)$, and there is a horizontal asymptote on the left or right at $y = 0$. Since it satisfies the horizontal line test, there is an inverse function: the base b logarithm, which is written $y = \log_b(x)$. The range is all of $(-\infty, \infty)$, and there's a vertical asymptote at $x = 0$. See **Figure 10**:

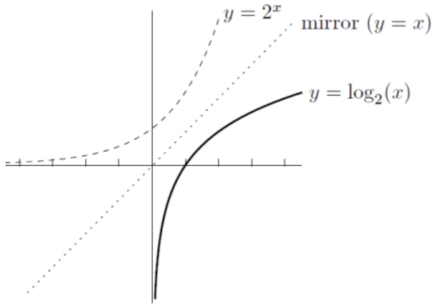


Figure 10. Base b is 2

CHAPTER 2 Review of Trigonometry

2.1. The Basics

- circumference; right-angled; hypotenuse; reciprocal;
- The circumference of a circle of radius 1 unit is 2π units, see **Figure 11**:

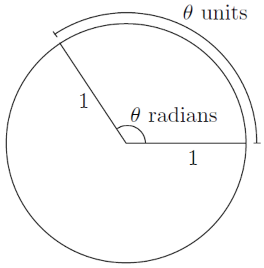


Figure 11. A circle of radius 1 unit

- The transfer formula between radians and degrees:

$$\mathit{angle\ in\ radians} = \frac{\pi}{180} \times \mathit{angle\ in\ degrees}$$

- A right-angled triangle **Figure 12**:

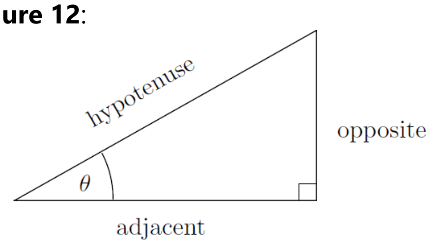


Figure 12. A right-angled triangle

$$\sin(\theta) = \frac{opposite}{hypotenuse}, \cos(\theta) = \frac{adjacent}{hypotenuse}, \text{and } \tan(\theta) = \frac{opposite}{adjacent}.$$

We'll also be using the reciprocal functions csc, sec, and cot:

$$\csc(x) = \frac{1}{\sin(x)}, \sec(x) = \frac{1}{\cos(x)}, \text{and } \cot(x) = \frac{1}{\tan(x)}.$$

- A nice table **Figure 13**:

	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
sin	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1
cos	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
tan	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	$\star$

Figure 13. A nice table (the star means that $tan(\pi/2)$ is undefined)

2.2. Extending the Domain of Trig Functions

2.2.1. The ASTC method

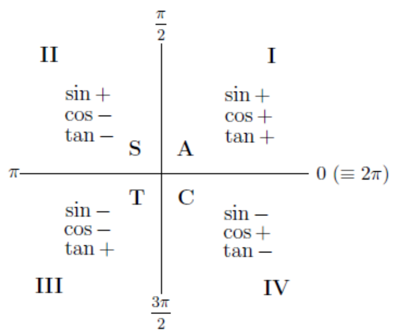


Figure 14. The ASTC method (All Sin Tan Cos)

2.3. The Graphs of Trig Functions

- $y = \sin(x)$ (period 2π, odd), see **Figure 15:**

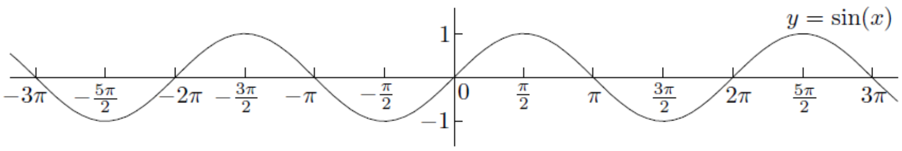


Figure 15. $y = sin(x)$

- $y = \cos(x)$ (period 2π, even), see **Figure 16:**

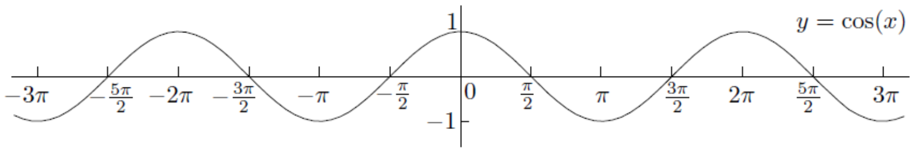


Figure 16. $y = cos(x)$

- $y = \tan(x)$ (period π, odd), see **Figure 17:**

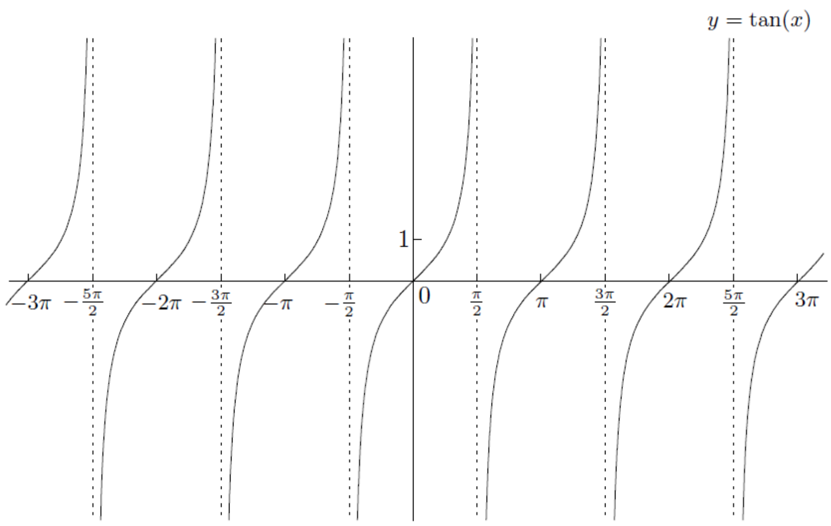


Figure 17. $y = tan(x)$

- $y = \sec(x)$ (period 2π, even), see **Figure 18:**

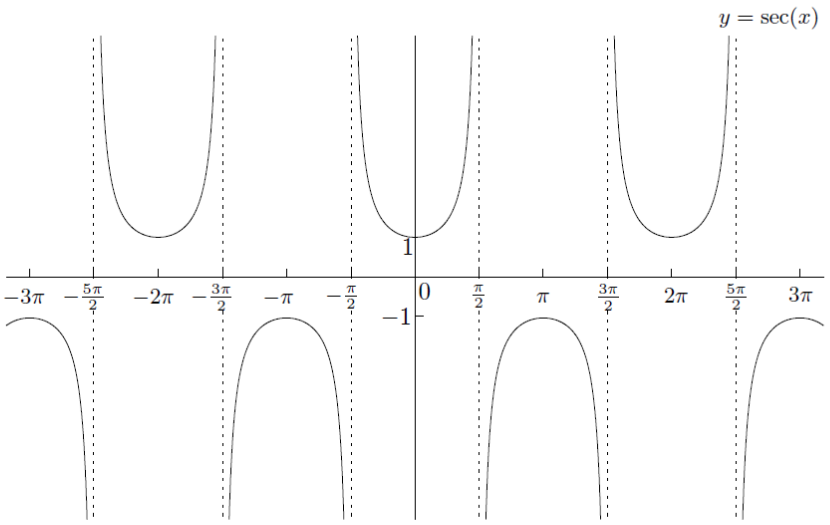


Figure 18. $y = sec(x)$

- $y = \csc(x)$ (period 2π, odd), see **Figure 19:**

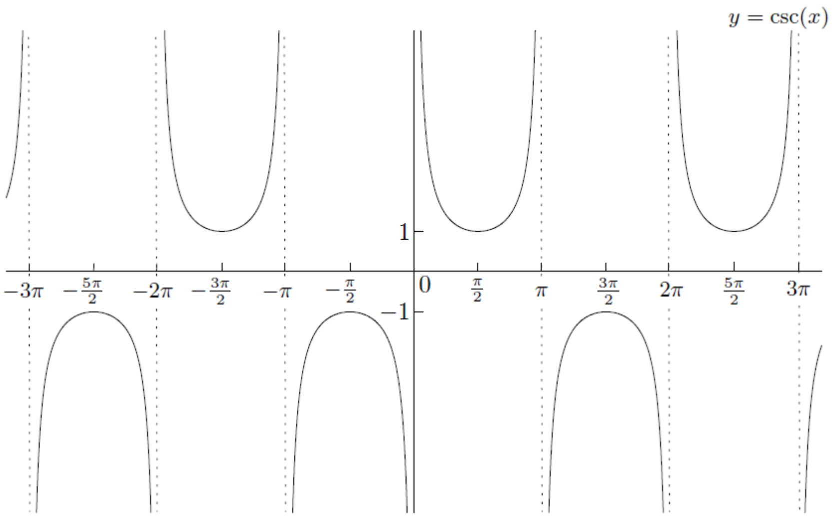


Figure 19. $y = csc(x)$

- $y = \cot(x)$ (period π, odd), see **Figure 20:**

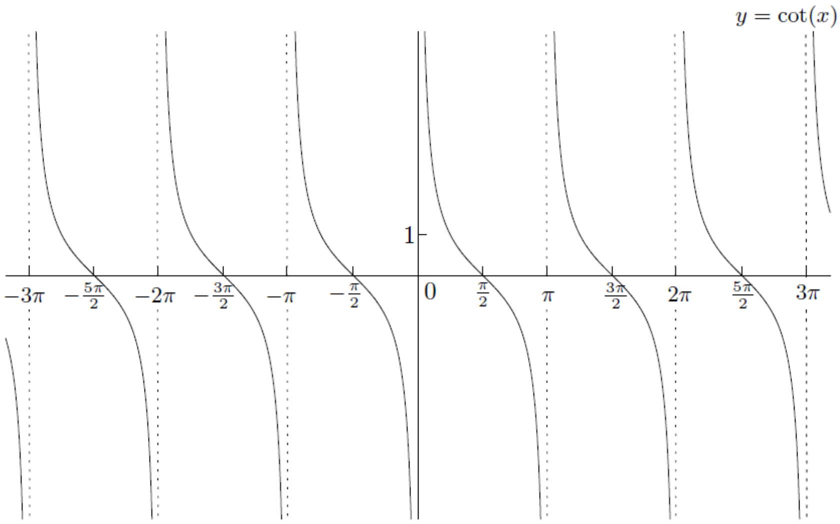


Figure 20. $y = cot(x)$

2.4. Trig Identities

$$\boxed{\cos^2(x) + \sin^2(x) = 1}$$

$$\boxed{1 + \tan^2(x) = \sec^2(x)}$$

$$\boxed{\cot^2(x) + 1 = \csc^2(x)}$$

● $trig\ function(x) = co - trig\ function(\frac{\pi}{2} - x)$:

$$\sin(x) = \cos(\frac{\pi}{2} - x)\,, \tan(x) = \cot(\frac{\pi}{2} - x)\,, \text{and } \sec(x) = \csc(\frac{\pi}{2} - x).$$

or

$$\cos(x) = \sin(\frac{\pi}{2} - x)\,, \cot(x) = \tan(\frac{\pi}{2} - x)\,, \text{and } \csc(x) = \sec(\frac{\pi}{2} - x).$$

Specifically, you should remember that

$$\boxed{\begin{aligned} \sin(A + B) &= \sin(A)\cos(B) + \cos(A)\sin(B) \\ \cos(A + B) &= \cos(A)\cos(B) - \sin(A)\sin(B) \end{aligned}}$$

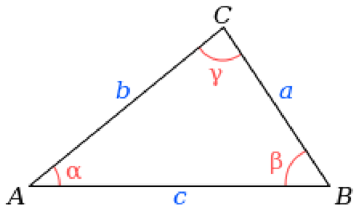
The double-angle formulas are

$$\boxed{\begin{gathered} \sin(2x) = 2\sin(x)\cos(x) \\ \cos(2x) = 2\cos^2(x) - 1 = 1 - 2\sin^2(x) \end{gathered}}$$

Supplement materials

Law of cosines: (Note it is useful for any triangle)

In **trigonometry**, the law of cosines (also known as the cosine formula, cosine rule, or al-Kashi's theorem) relates the lengths of the sides of a triangle to the cosine of one of its angles.

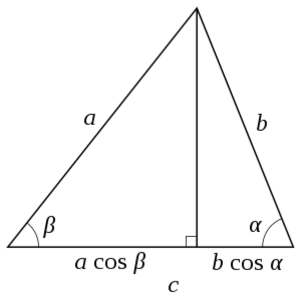


$$a^2 = b^2 + c^2 - 2bc\cos\alpha,$$
$$b^2 = a^2 + c^2 - 2ac\cos\beta,$$
$$c^2 = a^2 + b^2 - 2ab\cos\gamma.$$

The law of cosines **generalizes** the Pythagorean theorem, which holds only for right triangles. We can also get,

$$\cos\alpha = \frac{b^2 + c^2 - a^2}{2bc},$$
$$\cos\beta = \frac{a^2 + c^2 - b^2}{2ac},$$
$$\cos\gamma = \frac{a^2 + b^2 - c^2}{2ab}.$$

By **using** above equations, we can get the corresponding sine function. For example, below shows,



$$\cos\beta = \frac{a^2 + c^2 - b^2}{2ac}$$
$$\sin\beta = \sqrt{1 - \cos^2\beta}$$
$$\sin\beta = \frac{\text{Height}}{a} \rightarrow \text{Height} = a\sin\beta$$
$$\text{Area} = \frac{1}{2} \times c \times \text{Height} = \frac{1}{2}ac\sin\beta.$$

Observe this process, we can summary three useful equations,

$$\text{Area}_{\text{any triangle}} = \frac{1}{2}ac\sin\beta = \frac{1}{2}bc\sin\alpha = \frac{1}{2}ab\sin\gamma.$$

Also, by **using** above equations (Law of cosines), make some differences like below,

$$\cos\alpha \div \frac{1}{a} = \frac{b^2 + c^2 - a^2}{2abc},$$
$$\cos\beta \div \frac{1}{b} = \frac{a^2 + c^2 - b^2}{2abc},$$
$$\cos\gamma \div \frac{1}{c} = \frac{a^2 + b^2 - c^2}{2abc},$$

Next add them together,

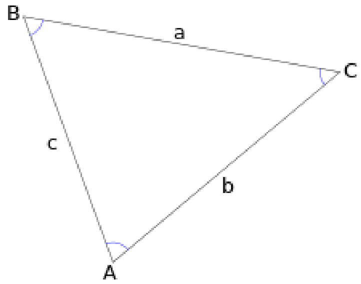
$$\frac{\cos\alpha}{a}+\frac{\cos\beta}{b}+\frac{\cos\gamma}{c}=\frac{b^2+c^2-a^2}{2abc}+\frac{a^2+c^2-b^2}{2abc}+\frac{a^2+b^2-c^2}{2abc}$$
$$=\frac{a^2+b^2+c^2}{2abc}.$$

So, this is a useful equation,

$$\frac{\cos\alpha}{a}+\frac{\cos\beta}{b}+\frac{\cos\gamma}{c}=\frac{a^2+b^2+c^2}{2abc}.$$

Law of sines: (Note it is useful for any triangle)

In **trigonometry**, the law of sines, sine law, sine formula, or sine rule is an equation relating the lengths of the sides of a triangle (any shape) to the sines of its angles. According to the law,



$$\frac{a}{\sin A}=\frac{b}{\sin B}=\frac{c}{\sin C}=d,$$

where a, b, and c are the lengths of the sides of a triangle, and A, B, and C are the opposite angles (see the figure above), while d is the diameter of the triangle's circumcircle.

3.1 Limits: The Basic Idea

● An example: $f(x) = x - 1 \quad \text{when } x \neq 2$

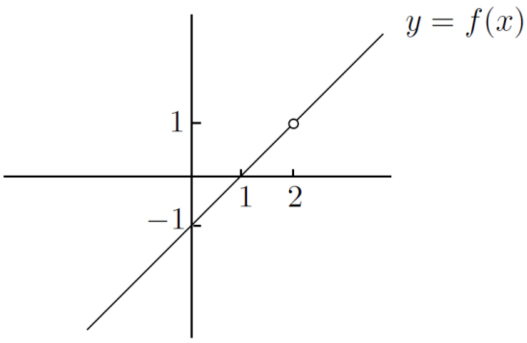


Figure 21. $\lim_{x\to 2} f(x) = 1$

● Modify it slightly: $g(x) = \begin{cases} x - 1 & \text{if } x \neq 2, \\ 3 & \text{if } x = 2. \end{cases}$

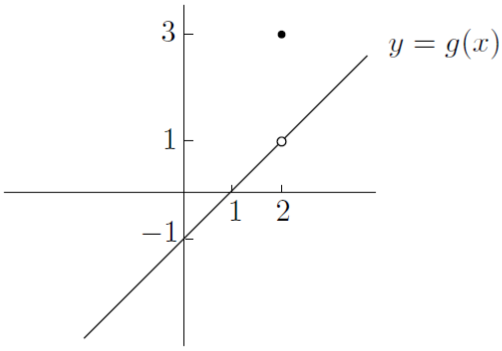
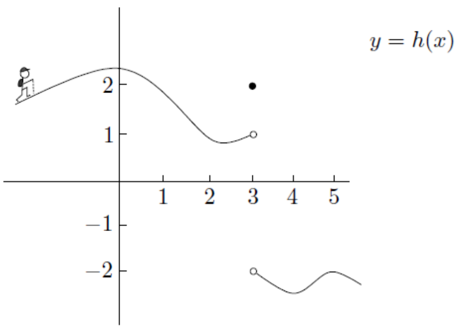


Figure 22. $\lim_{x\to 2} g(x) = 1$

3.2 Left-Hand and Right-Hand Limits

● How you would describe the behavior of $h(x)$ near $x = 3$:



We can summarize

$$\lim_{x\to 3^-} h(x) = 1 \qquad \text{and} \qquad \lim_{x\to 3^+} h(x) = -2$$

● The regular two-sided limit at $x = a$ exists **exactly when** both left-hand and right-hand limits at $x = a$ exist **and are equal to each other!**

I'm saying that

$$\lim_{x \to a^-} f(x) = L \qquad \text{and} \qquad \lim_{x \to a^+} f(x) = L$$

is the same thing as

$$\lim_{x \to a} f(x) = L$$

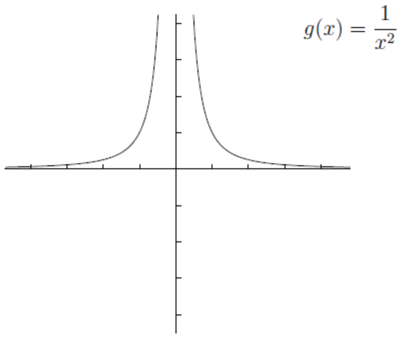
If the left-hand and right-hand limits are not equal, then the two-sided limit does not exist. We'd just write

$$\lim_{x \to 3} h(x) \ \text{does not exist}$$

3.3 When the Limit Does Not Exist

● A formal definition of the term "vertical asymptote":

"f has a vertical asymptote at $x = a$" means that at least one of $\lim_{x \to a^+} f(x)$ and $\lim_{x \to a^-} f(x)$ is equal to ∞ or $-\infty$.

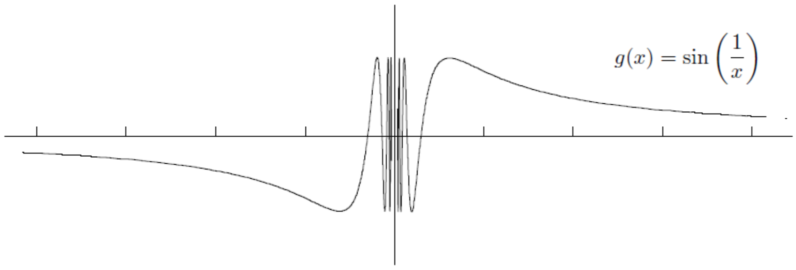


3.4 Limits at ∞ and $-\infty$

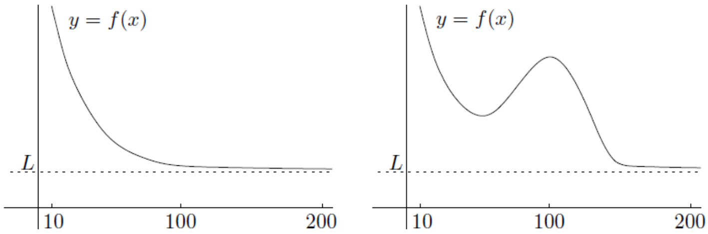
● How a function behaves when x gets really huge:

"f has a right-hand horizontal asymptote at $y = L$" means that $\lim_{x \to \infty} f(x) = L$.

"f has a left-hand horizontal asymptote at $y = M$" means that $\lim_{x \to -\infty} f(x) = M$.

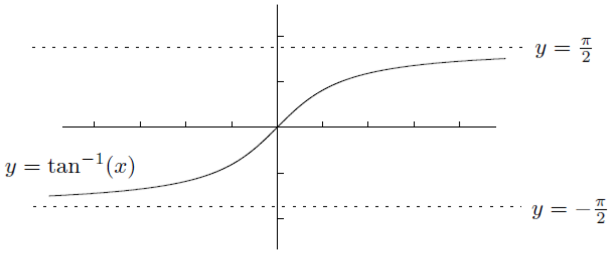


● Large numbers and small numbers



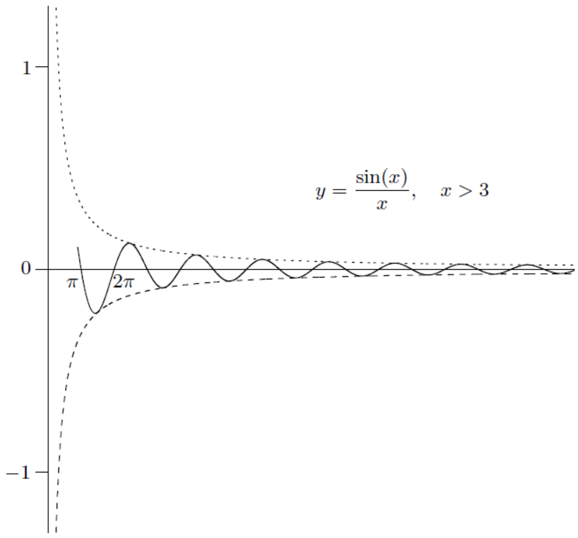
3.5 Two Common Misconceptions about Asymptotes

● First, a function doesn't have to have the same horizontal asymptote on the left as on the right



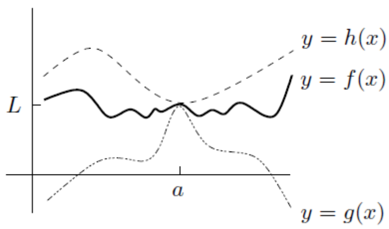
$$\lim_{x\to\infty} \tan^{-1}(x) = \frac{\pi}{2} \qquad \text{and} \qquad \lim_{x\to-\infty} \tan^{-1}(x) = -\frac{\pi}{2}$$

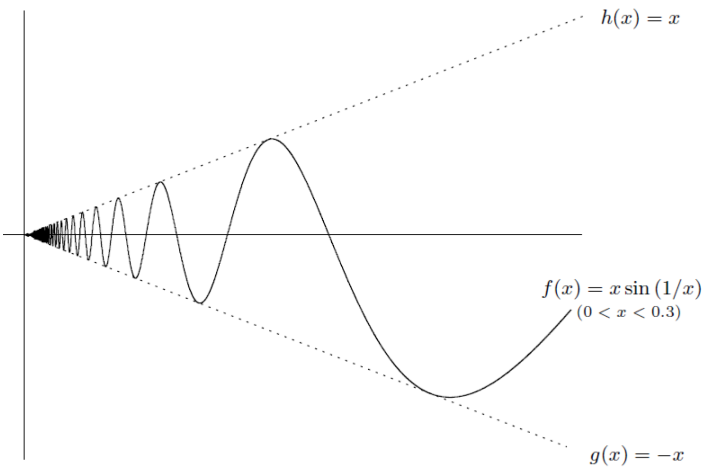
● A function can't cross its asymptote



3.6 The Sandwich Principle

● The sandwich principle, also known as the squeeze principle, says that if a function f is sandwiched between two functions g and h that converge to the same limit L as $x \to a$, then f also converges to L as $x \to a$.



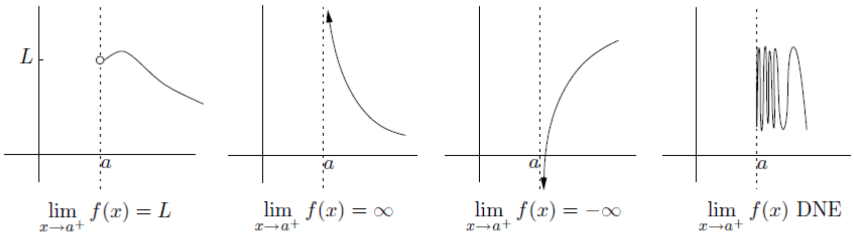


● In summary, here's what the sandwich principle says:

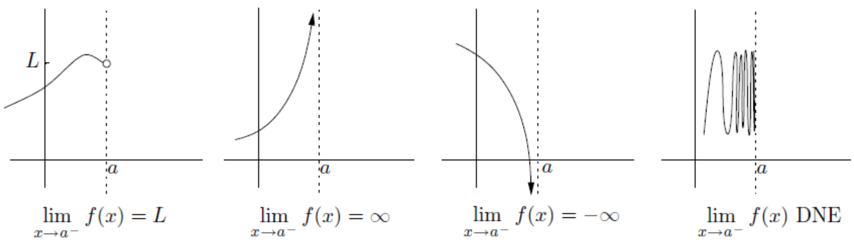
> If $g(x) \le f(x) \le h(x)$ for all x near a, and $\lim_{x\to a} g(x) = \lim_{x\to a} h(x) = L$, then $\lim_{x\to a} f(x) = L$.

3.7 Summary of Basic Types of Limits

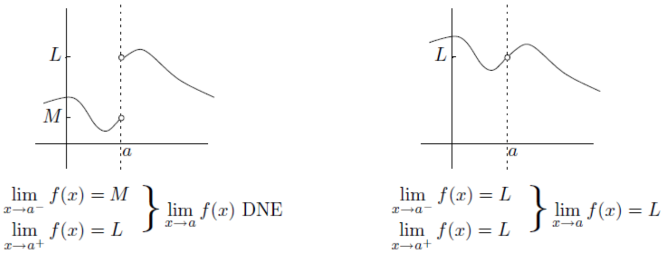
● The right-hand limit at $x = a$. Behavior of $f(x)$ to the left of $x = a$, and at $x = a$ itself, is irrelevant



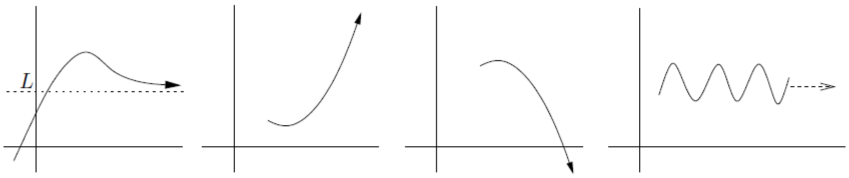
● The left-hand limit at $x = a$. Behavior of $f(x)$ to the right of $x = a$, and at $x = a$ itself, is irrelevant



● The two-sided limit at $x = a$

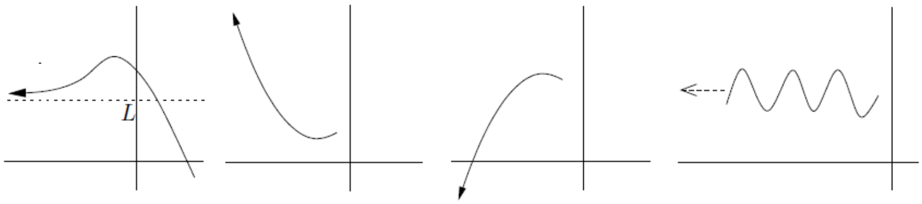


● The limit as $x \to \infty$



$\lim_{x \to \infty} f(x) = L$ $\quad$ $\lim_{x \to \infty} f(x) = \infty$ $\quad$ $\lim_{x \to \infty} f(x) = -\infty$ $\quad$ $\lim_{x \to \infty} f(x)$ DNE

- The limit as $x \to -\infty$



$\lim_{x \to -\infty} f(x) = L$ $\quad$ $\lim_{x \to -\infty} f(x) = \infty$ $\quad$ $\lim_{x \to -\infty} f(x) = -\infty$ $\quad$ $\lim_{x \to -\infty} f(x)$ DNE

CHAPTER 4 How to Solve Limit Problems Involving Polynomials

4.1 Limits Involving Rational Functions as $x \to a$

● Let's start off with limits that look like this:

$$\lim_{x\to a}\frac{p(x)}{q(x)}$$

where p and q are polynomials and a is a finite number. (Remember that the quotient $p(x)/q(x)$ of two polynomials is called a rational function.)

● The first thing you should always try is to substitute the value of a for x

● The formula for the difference of two cubes:

$$\boxed{a^3 - b^3 = (a-b)(a^2+ab+b^2)}$$

● Since we are taking limits, we use the plugging-in technique after factoring and canceling

4.2 Limits Involving Square Roots as $x \to a$

● Consider the following limit:

$$\lim_{x\to 5}\frac{\sqrt{x^2-9}-4}{x-5}$$

● conjugate expression: $(a-b)(a+b) = a^2 - b^2$

$$\lim_{x\to 5}\frac{\sqrt{x^2-9}-4}{x-5} = \lim_{x\to 5}\frac{\sqrt{x^2-9}-4}{x-5}\times\frac{\sqrt{x^2-9}+4}{\sqrt{x^2-9}+4}$$

$$\lim_{x\to 5}\frac{(x-5)(x+5)}{(x-5)\left(\sqrt{x^2-9}+4\right)} = \lim_{x\to 5}\frac{x+5}{\sqrt{x^2-9}+4}$$

4.3 Limits Involving Rational Functions as $x \to \infty$

● We are now trying to find limits of the form

$$\lim_{x\to\infty}\frac{p(x)}{q(x)}$$

● Here's a very important property of a polynomial: when x is large, the leading term dominates

● More generally, you can use the following theorem:

$$\boxed{\lim_{x\to\infty}\frac{C}{x^n} = 0}$$

for any $n > 0$, as long as C is constant.

● So we have proved that

$$\lim_{x\to\infty}\frac{p(x)}{\text{leading term of } p(x)} = 1$$

● Method and examples

$$\frac{p(x)}{\text{leading term of } p(x)} \times \left(\text{leading term of } p(x)\right)$$

4.4 Limits Involving Poly-type Functions as $x \to \infty$

● These aren't polynomials because they involve fractional powers or nth roots, for example, let's consider

$$\lim_{x\to\infty}\frac{\sqrt{16x^4+8}+3x}{2x^2+6x+1}$$

● Now let's see what happens when we modify the situation very slightly. Consider

$$\lim_{x\to\infty}\frac{\sqrt{16x^4+8}+3x^3}{2x^2+6x+1}$$

● But wait, you say-what if they are the same? For example, what is

$$\lim_{x\to\infty}\frac{\sqrt{4x^6-5x^5}-2x^3}{\sqrt[3]{27x^6+8x}}$$

Use conjugate expression.

4.5 Limits Involving Rational Functions as $x\to-\infty$

● Now let's spend a little time on limits of the form

$$\lim_{x\to-\infty}\frac{p(x)}{q(x)}$$

where p and q are polynomials or even poly-type functions. All the principles we've been using apply equally well here.

● There's only one other thing you have to beware

$\sqrt{x^2}=-x$ when x is negative

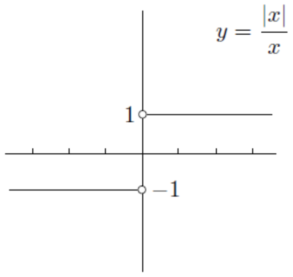
$\sqrt[3]{x^3}=x$ for all x (positive, negative, or zero)

If $x<0$ and you want to write $\sqrt[n]{x^{something}}=x^m$, the only thing you need a minus sign in front of x^m is when n is even and m is odd.

4.6 Limits Involving Absolute Values

● Sometimes you have to deal with functions involving absolute values. Consider this limit:

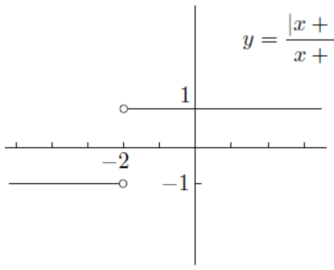
$$\lim_{x\to0^-}\frac{|x|}{x}$$



● A very slight variation of the above example is

$$\lim_{x\to(-2)^-}\frac{|x+2|}{x+2}$$

we see that it matters whether $x+2\geq0$ or $x+2\leq0$. These conditions can be rewritten as $x\geq-2$ or $x\leq-2$



CHAPTER 5 Continuity and Differentiability

5.1 Continuity

● The intuition is that you can draw the graph of the function in one piece, without lifting your pen off the page

● Continuity at a point

A function f is *continuous* as $x = a$ if $\lim_{x \to a} f(x) = f(a)$

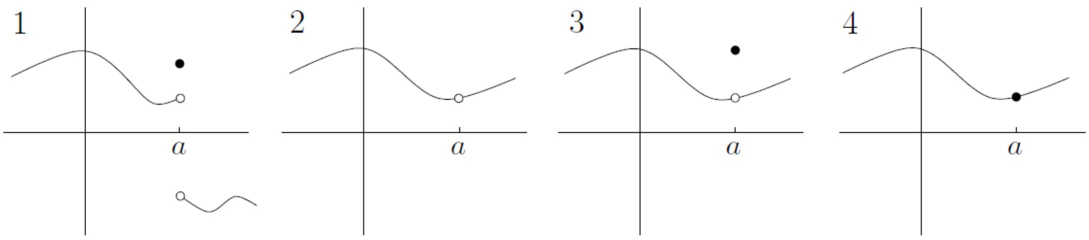
we can be a little more precise about the definition and explicitly require three things to be true:

1. The two-sided limit $\lim_{x \to a} f(x)$ exists (and is finite)

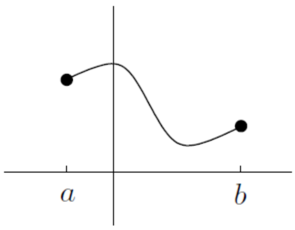
2. The function is defined at $x = a$; that is, $f(a)$ exists (and is finite)

3. The two above quantities are equal: that is,

$$\lim_{x \to a} f(x) = f(a)$$



● Continuity on an interval



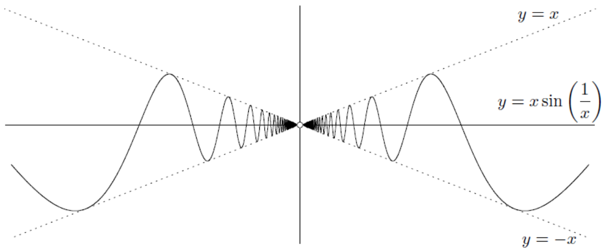
We say that a function f is continuous on $[a, b]$ if

1. the function f is continuous at every point in (a, b)

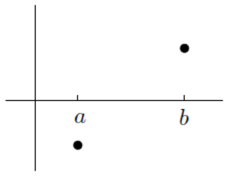
2. the function f is *right-continuous* at $x = a$. That is, $\lim_{x \to a^+} f(x)$ exists (and is finite), $f(a)$ exists, and these two quantities are equal; and

3. the function f is *left-continuous* at $x = b$. That is, $\lim_{x \to b^-} f(x)$ exists (and is finite), $f(b)$ exists, and these two quantities are equal.

● Examples of continuous functions. For example, every polynomial is continuous. Also, if you add, subtract, multiply or take the composition of two continuous functions, you get another continuous function. The same is almost true if you divide one continuous function by another: the quotient function is continuous everywhere except where the denominator is 0.



● Knowing that a function is continuous brings some benefits. The first is called the *Intermediate Value Theorem* (IVT)



Intermediate Value Theorem: if f is continuous on $[a, b]$, and $f(a) < 0$ and $f(b) > 0$, then there is at least one number c in the interval (a, b) such that $f(c) = 0$. The same is true if instead $f(a) > 0$ and $f(b) < 0$.

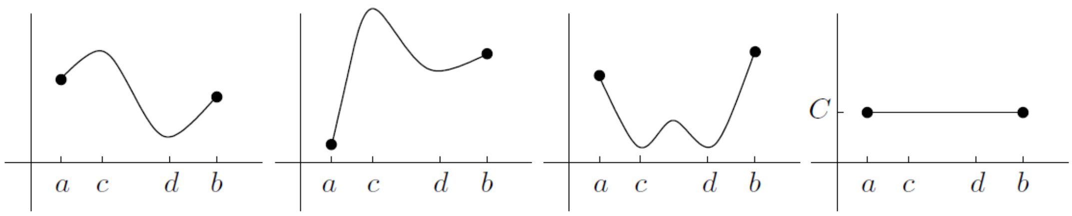
Here's a slightly harder example. How would you show that the equation $x = \cos(x)$ has a solution? The first step is to use a little trick: **put everything onto the left-hand side**. So, instead of solving $x = \cos(x)$, we try to solve $x - \cos(x) = 0$.

● A harder IVT example

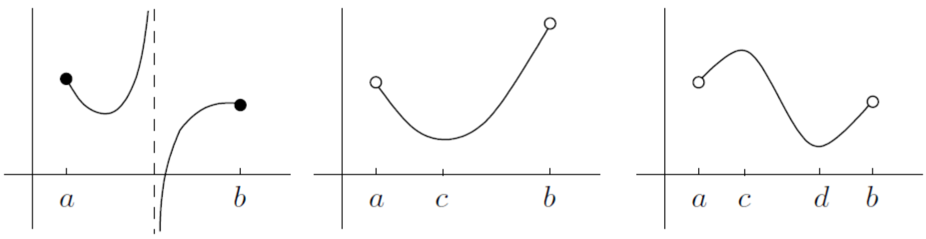
Any polynomial of odd degree has at least one root.

● The second benefit of knowing that a function is continuous: Maxima and minima of continuous functions

Max-Min Theorem: if f is continuous on $[a, b]$, then f has at least one maximum and one minimum on $[a, b]$.



If the function f isn't continuous, the following diagrams show some potential problems:



So, you can only use the theorem to guarantee the existence of a maximum and minimum in an interval $[a, b]$ if you know the function is continuous on the entire closed interval.

5.2 Differentiability

● Differentiability. This essentially means that the function has a derivative. One of the original inspirations for developing calculus came from trying to understand the relationship between speed, distance, and time for moving objects.

● Average speed

$$\text{speed} = \frac{\text{distance}}{\text{time}}$$

$$\text{displacement} = (\text{final position}) - (\text{initial position})$$

$$\text{average velocity} = \frac{\text{displacement}}{\text{time}}$$

● Instantaneous velocity

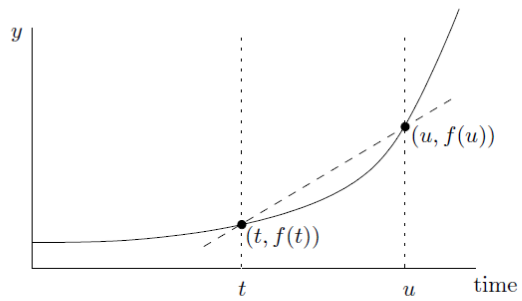
In particular, suppose that at time t, the car is at position $f(t)$. That is, let

$$f(t) = \text{position of car at time } t$$

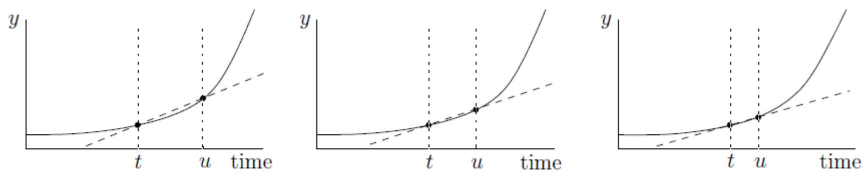
$$\text{instantaneous velocity at time } t = \lim_{h \to 0} \frac{f(t+h) - f(t)}{h}$$

● The graphical interpretation of velocity

We have a graphical interpretation for average velocity over the time period t to u



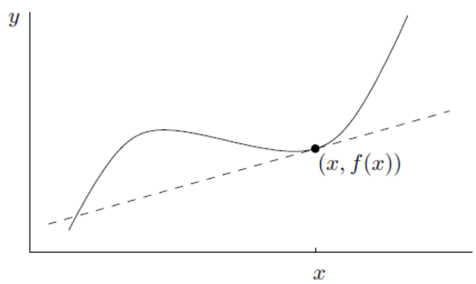
Try to find a similar interpretation for the instantaneous velocity



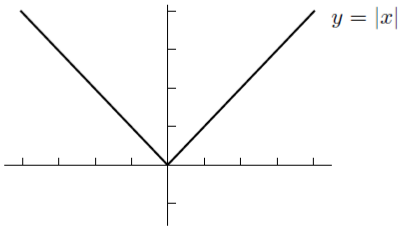
We'd like to say that the instantaneous velocity is exactly equal to the slope of the tangent line through $(t, f(t))$.

● Tangent lines

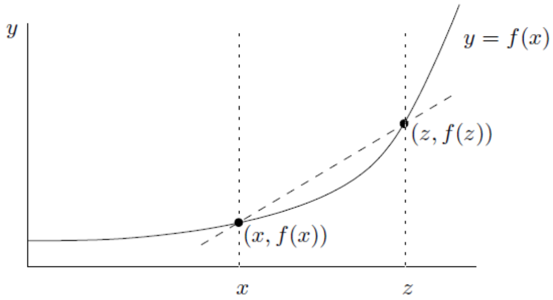
The tangent line doesn't have to intersect the curve only once!



It's possible that there's no tangent line through a given point on a graph

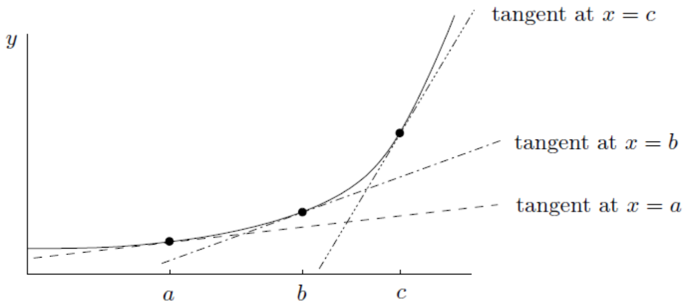


Start by picking a number z which is close to x



$$\text{slope of tangent line through } (x, f(x)) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \qquad \text{where } h = z - x$$

● The derivative function



$$\boxed{f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}}$$

As for the parabola $y = x^2$, its derivative is $f'(x) = 2x$.

● The derivative as a limiting ratio

Here the symbol Δ means "change in," so that Δx is just the change in x

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

Notice that $f'(x)$ isn't actually equal to the ratio of Δy to Δx: it's equal to the limit of that ratio as Δx tends to 0.

We'd now like to write dx, which should mean "really really tiny change in x," and similarly for y

If $y = f(x)$, then you can write $\frac{dy}{dx}$ instead of $f'(x)$

$\frac{dy}{dx}$ and $f'(x)$ are the same thing. Finally, remember that the quantity $\frac{dy}{dx}$ is not actually a fraction at all-it's the limit of the fraction $\frac{\Delta y}{\Delta x}$ as $\Delta x \to 0$.

● The derivative of linear functions

In our case, the graph of $y = mx + b$ is just a line of slope m and y-intercept equal to b. Then the tangent at any point on the line is just the line itself! This means that the value of $f'(x)$ should be m no matter what x is.

As you might expect, only linear functions have constant slope.

By the way, if f is actually constant, so that $f(x) = b$, then the slope is always 0. So we've proved that the derivative of a constant function is identically 0.

● Second and higher-order derivatives

There's a similar sort of notation for the second derivative:

If $y = f(x)$, then you can write $\frac{d^2y}{dx^2}$ instead of $f''(x)$

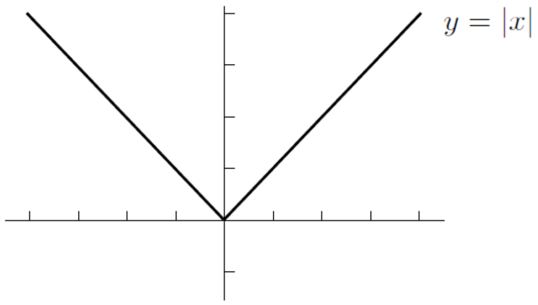
The third derivative of f as being the derivative of the second derivative of f

$$f'''(x),\quad f^{(3)}(x),\quad \frac{d^3y}{dx^3},\quad \text{or}\ \frac{d^3}{dx^3}(y)$$

The notation $f^{(3)}(x)$ is particularly convenient for higher derivatives. That way, any derivative can be written in the form $f^{(n)}(x)$ for some integer n.

● When the derivative does not exist

We know the graph of $f(x) = |x|$ has a sharp corner at the origin



The point is that if $f(x) = |x|$, at $x = 0$ the right-hand derivative is 1 and the left-hand derivative is -1. Since the left-hand slope doesn't equal the right-hand slope, there can be no derivative at $x = 0$.

In conclusion, there are continuous functions which are not differentiable, but all differentiable functions are continuous.

● Differentiability and continuity

If a function f is differentiable at x, then it's continuous at x.

Differentiable functions are automatically continuous. Remember, though, that continuous functions aren't always differentiable!

So, how do we prove our big claim? Let's start by seeing what we want to prove. To show that f is continuous at x, we're going to need to show that

$$\lim_{u \to x} f(u) = f(x)$$

Before we proceed farther, I want to substitute $h = u - x$ as we've done before. In that case, $u = x + h$, and as $u \to x$, we see that $h \to 0$. So the above equation can be replaced by

$$\lim_{h\to 0} f(x+h) = f(x)$$

Now that we are aware of our destination, let's start with what we actually know. Well, we know that f is differentiable at x; this means that $f'(x)$ exists, so by the definition of f', the limit

$$\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

exists. We know that $f(x)$ exists. We still need to do something clever. The trick is to start with another limit:

$$\lim_{h\to 0} \left(\frac{f(x+h)-f(x)}{h} \times h\right)$$

we can work out this limit exactly by splitting it into two factors:

$$\lim_{h\to 0} \left(\frac{f(x+h)-f(x)}{h} \times h\right) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \times \lim_{h\to 0} h = f'(x) \times 0 = 0$$

On the other hand, we could have taken the original limit and instead canceled out the factor of h to get

$$\lim_{h\to 0} \left(\frac{f(x+h)-f(x)}{h} \times h\right) = \lim_{h\to 0} \big(f(x+h) - f(x)\big)$$

Of course, the value of $f(x)$ doesn't depend on the limit at all:

$$\left(\lim_{h\to 0} f(x+h)\right) - f(x) = 0$$

$$\lim_{h\to 0} f(x+h) = f(x)$$

CHAPTER 6 How to Solve Differentiation Problems

6.1 Finding Derivatives Using the Definition

$$\boxed{\frac{d}{dx}\left(\frac{1}{x}\right) = -\frac{1}{x^2}}$$

$$\boxed{\frac{d}{dx}(\sqrt{x}) = \frac{1}{2\sqrt{x}}}$$

$$\boxed{\frac{d}{dx}(x^a) = ax^{a-1}}$$

If C is constant, then $\frac{d}{dx}(C) = 0$

$$\boxed{\frac{d}{dx}(x) = 1}$$

6.2 Finding Derivatives (the Nice Way)

- Constant multiples of functions

$$\frac{d}{dx}(7x^2) = 7 \times 2x$$

- Sums and differences of functions

It's even easier to differentiate sums and differences of functions: just differentiate each piece and then add or subtract.

- Products of functions via the product rule

Product rule (version 1): if $h(x) = f(x)g(x)$, then $h'(x) = f'(x)g(x) + f(x)g'(x)$.

Product rule (version 2): if $y = uv$, then $\frac{dy}{dx} = v\frac{du}{dx} + u\frac{dv}{dx}$.

Product rule (three variables): if $y = uvw$, then $\frac{dy}{dx} = \frac{du}{dx}vw + u\frac{dv}{dx}w + uv\frac{dw}{dx}$.

- Quotients of functions via the quotient rule

Quotients rule (version 1): if $h(x) = \frac{f(x)}{g(x)}$, then $h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{\left(g(x)\right)^2}$.

Quotients rule (version 2): if $y = \frac{u}{v}$, then $\frac{dy}{dx} = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$.

- Composition of functions via the chain rule

Chain rule (version 1): if $h(x) = f\left(g(x)\right)$, then $h'(x) = f'\left(g(x)\right)g'(x)$.

Chain rule (version 2): if y is a function of u, and u is a function of x , then $\frac{dy}{dx}=\frac{dy}{du}\frac{du}{dx}$.

● A nasty example

● Justification of the product rule and the chain rule

6.3 Finding the Equation of a Tangent Line

● One benefit of finding derivatives is that you can use derivatives to find the equation of a tangent line to a given curve

1. find the slope
2. find a point on the line (x_0, y_0)
3. use the point-slope form $y - y_0 = m(x - x_0)$.

6.4 Velocity and Acceleration

● Another application of finding derivatives is to compute velocities and accelerations of moving objects

$$\text{velocity } = v = \frac{dx}{dt}$$

$$\text{acceleration } = a = \frac{dv}{dt} = \frac{d^2x}{dt^2}$$

● Constant negative acceleration

An object thrown at time $t = 0$ from initial height h with initial velocity u satisfies the equations

$$a = -g, \quad v = -gt + u, \quad \text{and} \quad x = -\frac{1}{2}gt^2 + ut + h$$

6.5 Limits Which Are Derivatives in Disguise

$$\lim_{h\to 0}\frac{\sqrt[5]{32+h}-2}{h} \quad \left(\lim_{h\to 0}\frac{\sqrt[5]{x+h}-\sqrt[5]{x}}{h} = \frac{1}{5}x^{-4/5}\right)$$

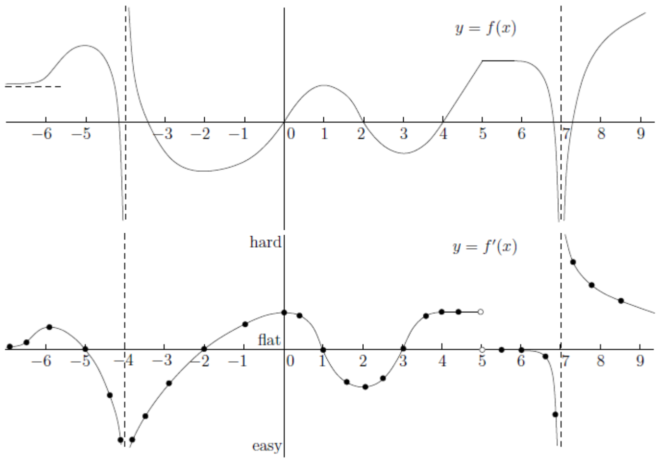
$$\lim_{h\to 0}\frac{\sqrt{(4+h)^3-7(4+h)}-6}{h} \quad \left(\lim_{h\to 0}\frac{\sqrt{(x+h)^3-7(x+h)}-\sqrt{x^3-7x}}{h} = \frac{3x^2-7}{2\sqrt{x^3-7x}}\right)$$

If you get stuck on a limit, it might be a derivative in disguise. Telltale signs are that the dummy variable is by itself in the denominator, and the numerator is the difference of two quantities. Even if this doesn't happen, you could still be dealing with a derivative in disguise

$$\lim_{h\to 0}\frac{h}{(x+h)^6-x^6} \quad \left(\text{flip it: } \lim_{h\to 0}\frac{(x+h)^6-x^6}{h} = 6x^5\right)$$

6.6 Derivatives of Piecewise-Defined Functions

6.7 Sketching Derivative Graphs Directly



y = f(x)
−6
−5
−3
−2
−1
0
1
2
3
4
5
6
7
8
9
hard
y = f′(x)
flat
easy

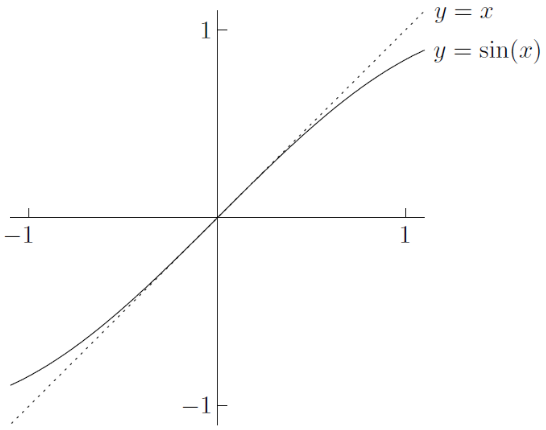
7.1 Limits Involving Trig Functions

$$\lim_{x\to 0}\frac{\sin(5x)}{x} \quad \text{and} \quad \lim_{x\to\infty}\frac{\sin(5x)}{x}$$

● The small case

It is true in the limit as $x \to 0$:

$$\boxed{\lim_{x\to 0}\frac{\sin(x)}{x} = 1}$$



Look at $\tan(x)$

$$\lim_{x\to 0}\frac{\tan(x)}{x} = \lim_{x\to 0}\frac{\frac{\sin(x)}{\cos(x)}}{x} = \lim_{x\to 0}\left(\frac{\sin(x)}{x}\right)\left(\frac{1}{\cos(x)}\right) = (1)\left(\frac{1}{1}\right) = 1$$

So we have shown that

$$\boxed{\lim_{x\to 0}\frac{\tan(x)}{x} = 1}$$

What happens to $\cos(x)/x$ as $x \to 0$

$$\lim_{x\to 0^+}\frac{\cos(x)}{x} = \infty, \quad \lim_{x\to 0^-}\frac{\cos(x)}{x} = -\infty, \quad \text{so} \quad \lim_{x\to 0}\frac{\cos(x)}{x} \ \text{DNE}$$

● Solving problems-the small case

$$\lim_{x\to 0}\frac{\sin(x^2)}{x^2} = 1$$

$$\boxed{\lim_{x\to 0}\frac{\sin(\text{small})}{\text{same small}} = 1} \quad \text{and} \quad \boxed{\lim_{x\to 0}\frac{\tan(\text{small})}{\text{same small}} = 1}$$

$$\boxed{\lim_{x\to 0}\cos(\text{small}) = 1}$$

$$\boxed{\lim_{x\to 0}\frac{1-\cos(x)}{x} = 0}$$

● The large case

$$\boxed{-1 \leq \sin(x) \leq 1} \quad \text{and} \quad \boxed{-1 \leq \cos(x) \leq 1} \quad \text{for any } x$$

Using the sandwich principle, you can treat $\sin(\text{anything})$ or $\cos(\text{anything})$ as being of lower degree than any positive power of x, so long as you are only adding or subtracting.
More precisely, if you are solving a problem of the form

$$\lim_{x\to\infty}\frac{p(x)}{q(x)}$$

where p and q are polynomials or poly-type functions but with some sines and cosines added on, then the degrees of the top and bottom are the same as they would be without the sines and cosines added on. The only exception is when p or q has degree 0; then the trig part could be significant.
In practice, most mathematicians would have established the general principle that

$$\lim_{x\to\infty}\frac{\sin(\text{anything})}{x^{\alpha}}=0$$

for any positive exponent α, and similarly when sine is replaced by cosine.

● The "other" case
Consider the limit

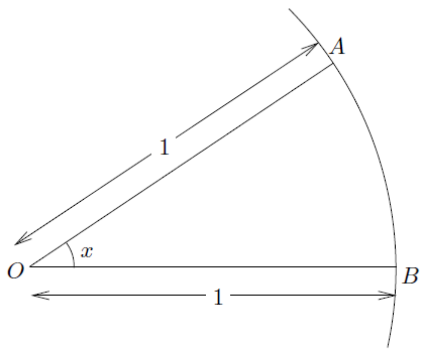
$$\lim_{x\to\pi/2}\frac{\cos(x)}{x-\frac{\pi}{2}}$$

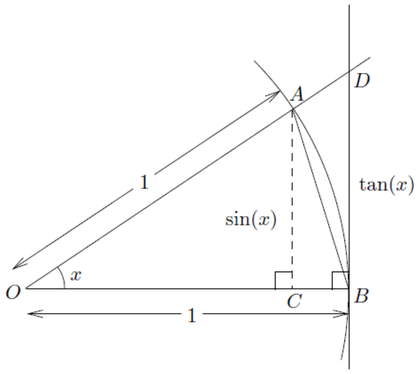
A good general principle when dealing with a limit involving $x\to a$ for some $a\neq 0$ is to **shift the problem to 0 by substituting** $t=x-a$

$$\lim_{x\to\pi/2}\frac{\cos(x)}{x-\frac{\pi}{2}}=\lim_{t\to 0}\frac{\cos\left(t+\frac{\pi}{2}\right)}{t}=\lim_{t\to 0}\frac{-\sin(t)}{t}$$

● Proof of an important limit
Now it's time to prove it

$$\lim_{x\to 0}\frac{\sin(x)}{x}=1$$





$$\boxed{\sin(x) < x < \tan(x) \qquad \text{for } 0 < x < \frac{\pi}{2}}$$

Let's first take reciprocals of the nice inequality, and multiply by the positive quantity $\sin(x)$

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

By the sandwich principle

$$\lim_{x \to 0^+} \frac{\sin(x)}{x} = 1$$

To prove that the left-hand limit is 1 set $t = -x$.

7.2 Derivatives Involving Trig Functions

$$\boxed{\frac{d}{dx}\sin(x) = \cos(x)}$$

$$\boxed{\frac{d}{dx}\cos(x) = -\sin(x)}$$

$$\boxed{\frac{d}{dx}\tan(x) = \sec^2(x)}$$

$$\boxed{\frac{d}{dx}\sec(x) = \sec(x)\tan(x)}$$

$$\boxed{\frac{d}{dx}\csc(x) = -\csc(x)\cot(x)}$$

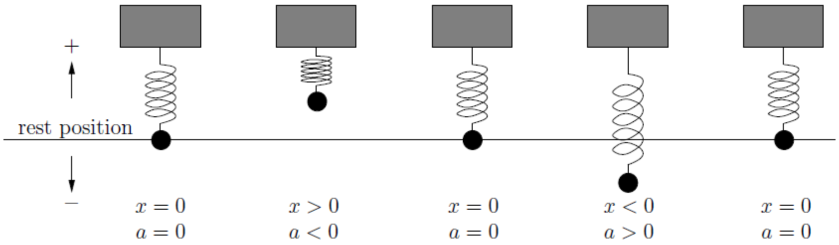
$$\boxed{\frac{d}{dx}\cot(x) = -\csc^2(x)}$$

- Examples of differentiating trig functions
- Simple harmonic motion

$$x = 3\sin(4t)$$

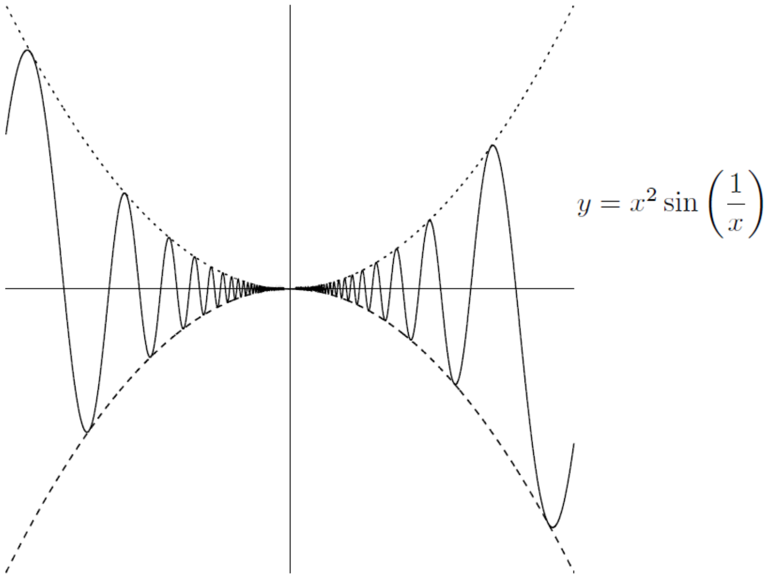
$$v = \frac{d}{dt}(3\sin(4t)) = 12\cos(4t)$$

$$a = \frac{dv}{dt} = -48\cos(4t)$$



● A curious function

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$



$$f'^{(0)} = 0$$

But $\lim_{x\to 0^+} f'(x) = \lim_{x\to 0^+}\left(2x\sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right)\right)$ DNE, neither does $\lim_{x\to 0^-} f'(x)$.

So, there are functions out there which are differentiable, yet their derivatives aren't continuous.

CHAPTER 8 Implicit Differentiation and Related Rates

8.1 Implicit Differentiation

Consider the following two derivatives:

$$\frac{d}{dx}(x^2) \qquad \text{and} \qquad \frac{d}{dx}(y^2)$$

The best way is to say to yourself that the first of the derivatives above is asking how much the quantity x^2 changes when we change x a little bit. On the other hand, think of second this way, if you change x, then y will change a little bit; this change in y will cause y^2 to change. (All this is true only if y depends on x, of course-if not, then when you change x, nothing at all will happen to y.

● Techniques and examples

Now it's time to get practical. Consider the following equation:

$$x^2 + y^2 = 4$$

whack a d/dx in front of both sides:

$$\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}(4)$$

$$\frac{d}{dx}(x^2) + \frac{d}{dx}(y^2) = \frac{d}{dx}(4)$$

$$2x + 2y\frac{dy}{dx} = 0$$

$$\frac{dy}{dx} = -\frac{x}{y}$$

Here's another example: if

$$5\sin(x) + 3\sec(y) = y - x^2 + 3$$

Now let's differentiate the above equation:

$$\frac{d}{dx}(5\sin(x)) + \frac{d}{dx}(3\sec(y)) = \frac{dy}{dx} - \frac{d}{dx}(x^2) + \frac{d}{dx}(3)$$

$$5\cos(x) + 3\sec(y)\tan(y)\frac{dy}{dx} = \frac{dy}{dx} - 2x$$

$$\frac{dy}{dx} = \frac{2x + 5\cos(x)}{1 - 3\sec(y)\tan(y)}$$

Finally, plug in $x = 0$ and $y = 0$ to see that:

$$\frac{dy}{dx} = \frac{2(0) + 5\cos(0)}{1 - 3\sec(0)\tan(0)} = \frac{2(0) + 5(1)}{1 - 2(1)(0)} = 5$$

But do you see how we might have saved a little effort? Go back to the second equation, we could have saved a little time by plugging $x = 0$ and $y = 0$ into the above equation, which can easily reduce $dy/dx = 5$. So a good rule of thumb is **that if you only need the derivative at a certain point, substitute before rearranging**-it often saves time.

Here's a brief summary of the above methods:

1. in your original equation, differentiate everything and simplify using the chain, product, and quotient rules;
2. if you want to find dy/dx, rearrange and divide to solve for dy/dx; but
3. if instead you want to find the slope or equation of the tangent at a particular point on the curve, first substitute the known values of x and y, then rearrange to find dy/dx. Then use

the point-slope formula to find the equation of the tangent, if needed.

● Finding the second derivative implicitly

It's also possible to differentiate twice to get the second derivative. For example, if

$$2y+\sin(y)=\frac{x^2}{\pi}+1$$

Now, if you want to differentiate twice, you have to start by differentiating once! You should get

$$2\frac{dy}{dx}+\cos(y)\frac{dy}{dx}=\frac{2x}{\pi}$$

Differentiate the above equation with respect to x:

$$\frac{d}{dx}\left(2\frac{dy}{dx}\right)+\frac{d}{dx}\left(\cos(y)\frac{dy}{dx}\right)=\frac{d}{dx}\left(\frac{2x}{\pi}\right)$$

Beware: the quantities

$$\left(\frac{dy}{dx}\right)^2 \quad \text{and} \quad \frac{d^2y}{dx^2}$$

are completely different!

Finally, we can write this as

$$2\frac{d^2y}{dx^2}-\sin(y)\left(\frac{dy}{dx}\right)^2+\cos(y)\frac{d^2y}{dx^2}=\frac{2}{\pi}$$

Phew. That was exhausting. We're not done yet, though: we still need to find d^2y/dx^2 when $x=\pi$ and $y=\pi/2$. So plug that in to the above equation: you get

$$2\frac{d^2y}{dx^2}-\sin\left(\frac{\pi}{2}\right)\left(\frac{dy}{dx}\right)^2+\cos\left(\frac{\pi}{2}\right)\frac{d^2y}{dx^2}=\frac{2}{\pi}$$

We still need dy/dx! So put $x=\pi$ and $y=\pi/2$ to second equation above and get

$$\frac{dy}{dx}=1$$

Put that into our second derivative equation and we will get d^2y/dx^2

$$\frac{d^2y}{dx^2}=\frac{1}{\pi}+\frac{1}{2}$$

when $x=\pi$ and $y=\pi/2$, so we're finally done!

8.2 Related Rates

Consider two quantities-they can measure anything you like-that are related to each other. If you know one, you can find the other.

Of course, as one of the two quantities changes, so does the other. Suppose that we know how fast one of the quantities is changing. Then how fast is the other one changing? That is exactly what we mean by the term *related rates*. You see, a *rate of change* is the speed at which a quantity is changing over time.

Here's the real definition: **the rate of change of a quantity Q is the derivative of Q with respect to time**. That is,

if Q is some quantity, then the rate of change of Q is $\frac{dQ}{dt}$.

When you see the word “rate,” you should automatically think “d/dt.”

So, let's look at a general overview of how to solve problems involving related rates:

1. Read the question. Identify all the quantities and note which one you need to find the rate of. Draw a picture if you need to!
2. Write down an equation (sometimes you need more than one) that relates all the quantities. To do this, you may need to do some geometry, possibly involving similar triangles. If you have more than one equation, try to solve them simultaneously to eliminate unnecessary variables.
3. Differentiate your remaining equation(s) implicitly with respect to time t. That is, whack both sides of each equation with a $\frac{d}{dt}$. You end up with one or more equations relating the rates of change.
4. Finally, substitute values for everything you know into all the equations you have. Solve the equations simultaneously to find the rate you need.

Just one more thing before we look at examples: it's vital that you **substitute values at the end, after differentiating**! That is, don't switch steps 3 and 4. If you substitute values first, denying the quantities the ability to change, then your rates will all be 0. That's what you get for freezing everything in place.

- A simple example

Suppose that a perfectly spherical balloon is being inflated by a pump. Air is entering the balloon at the constant rate of 12π cubic inches per second. At what rate does the radius of the balloon change at the instant when the radius itself is 2 inches? Also, at what rate does the radius change when the volume is 36π cubic inches?

OK, let's write down our quantities (step 1):

These are the volume and the radius of the balloon. Let's call the volume V (in cubic inches) and the radius r (in inches).

Now, we need an equation relating V and r (step 2):

Here's where the geometry comes in. Since the balloon is a sphere, we know that

$$V = \frac{4}{3}\pi r^3$$

Now we need to relate the rates (step 3):

Differentiate both sides implicitly with respect to t:

$$\frac{d}{dt}(V) = \frac{d}{dt}\left(\frac{4}{3}\pi r^3\right) = 4\pi r^2 \frac{dr}{dt}$$

Finally, we're ready to substitute (step 4):

In symbols, we have $dV/dt = 12\pi$, Plugging this into the above equation, we get

$$12\pi = 4\pi r^2 \frac{dr}{dt}$$

Rearranging leads to

$$\frac{dr}{dt} = \frac{3}{r^2}$$

Armed with the formula, we can quickly do both parts of the question. In the first part, we know that the radius is 2 inches, so set $r = 2$ in our formula from above:

$$\frac{dr}{dt} = \frac{3}{2^2} = \frac{3}{4}$$

So the answer is $\frac{3}{4}$. But $\frac{3}{4}$ what? It's important to write a sentence summarizing the situation, as well as including the **units** of measurement. In this case, we'd say that when the radius is 2 inches, the rate of change of the radius is $\frac{3}{4}$ inches per second.

Now, for the second part of the question, we know that the volume is 36π cubic inches. Put $V = 36\pi$ and solve for r, you should be able to see that $r = 3$ inches. Finally, substituting into the equation for dr/dt gives

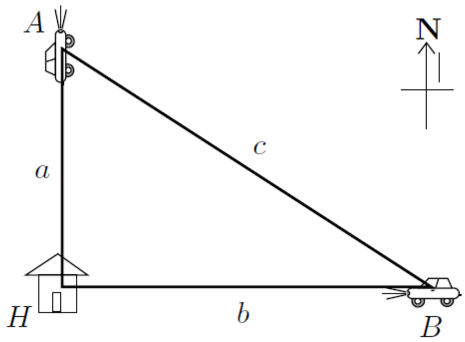
$$\frac{dr}{dt} = \frac{3}{r^2} = \frac{1}{3}$$

So when the volume is 36π cubic inches, the rate of change of the radius is $\frac{1}{3}$ inches per second.

● A slightly harder example

Let's look at another relatively straightforward example, this time involving three quantities. Suppose there are two cars, A and B. Car A is driving on a road heading directly north away from your house, and car B is driving on a different road heading directly west toward your house. Car A travels at 55 miles per hour and car B travels at 45 miles per hour. At what rate is the distance between the cars changing when A is 21 miles north of your house and car B is 28 miles east of your house?

To answer this question, we'd better draw a picture (step 1): Draw your house H and the cars A and B. Let the distance between H and A be given by a; let the distance between H and B be called b; and let the distance between the cars be called c. The diagram looks like this:



Time for step 2. The equation relating a, b, and c is nothing other than Pythagoras' Theorem:

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

Moving on to step 3, we differentiate implicitly with respect to time t. Make sure you agree that we get

$$2a\frac{da}{dt} + 2b\frac{db}{dt} = 2c\frac{dc}{dt}$$

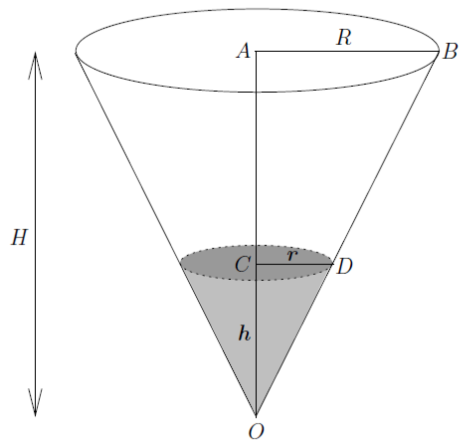
Now, we know that car A is moving at 55 miles an hour away from your house. This means that the distance a is increasing by 55 miles per hour, so $da/dt = 55$. As for B, it is moving at 45 miles an hour toward your house. This means that b is decreasing by 45 miles an hour, so $db/dt = -45$

The end result is that $dc/dt = -3$.

● A much harder example

Here's a tougher example involving similar triangles: suppose there's a freakin' huge water tank in the shape of a cone (with the point at the bottom). The height of the cone is twice the radius of the cone. Water is being pumped into the tank at the rate of 8π cubic feet per second. At what rate is the water level changing when the volume of water in the tank is 18π cubic feet? There's a second part as well: assume that the tank develops a little hole at the bottom that causes water to flow out at a rate of one cubic foot per second for every cubic foot of water in the tank. I want to know the same thing as before: at what rate is the water level changing when the volume of water in the tank is 18π cubic feet, but now with the leak in the tank?

Let's start with the first part. Here's a diagram of the situation:



The height of the tank is H and its radius is R. The height of the water level is h and the radius of the top of the water surface is r. All these quantities are measured in feet. Let's also let v be the volume of water in the tank, measured in cubic feet (step 1)

For step 2, we have to start relating some of these quantities. We are given that the tank's height is twice the radius, so we have $H = 2R$. There are some similar triangles in the diagram: in fact, ΔABO is similar to ΔCDO, so $H/R = h/r$. Since $H = 2R$, we have $2R/R = h/r$, which means that $h = 2r$. The volume of a cone of height h units and radius r units is given by $v = \frac{1}{3}\pi r^2 h$ cubic units. Using the equation $r = h/2$, we have

$$v = \frac{1}{3}\pi r^2 h = \frac{1}{3}\pi\left(\frac{h}{2}\right)^2 h = \frac{\pi h^3}{12}$$

Now, for step 3, let's differentiate this with respect to time t. By the chain rule,

$$\frac{dv}{dt} = \frac{\pi}{12} \times 3h^2 \frac{dh}{dt} = \frac{\pi h^2}{4}\frac{dh}{dt}$$

Great-now for step 4, substitute in everything we know into the two equations

Above. We know that $dv/dt = 8\pi$ and we're interested in what happens when $v = 18\pi$. Substituting, we get

$$18\pi = \frac{\pi h^3}{12} \qquad \text{and} \qquad 8\pi = \frac{\pi h^2}{4}\frac{dh}{dt}$$

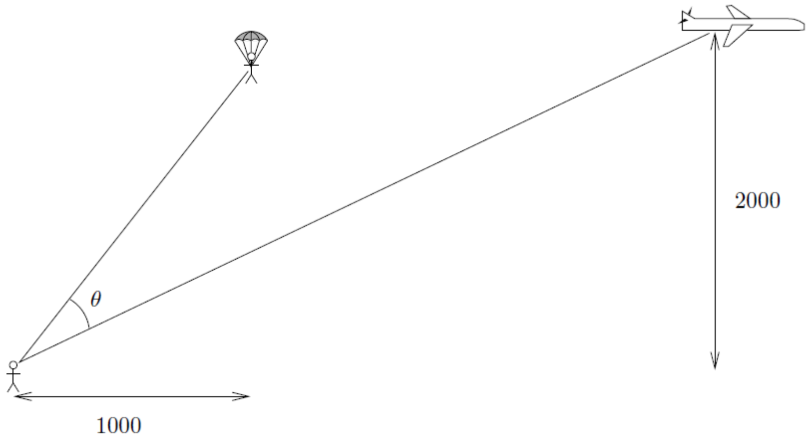
Final, we get $dh/dt = 8/9$.

The second part is almost the same. In fact, the only difference occurs at step 4. We know one cubic foot is leaving per second for every cubic foot of water in the tank. Since there are v cubic feet of water in the tank (by definition!), the rate of outflow from the leak is v cubic feet per second. So

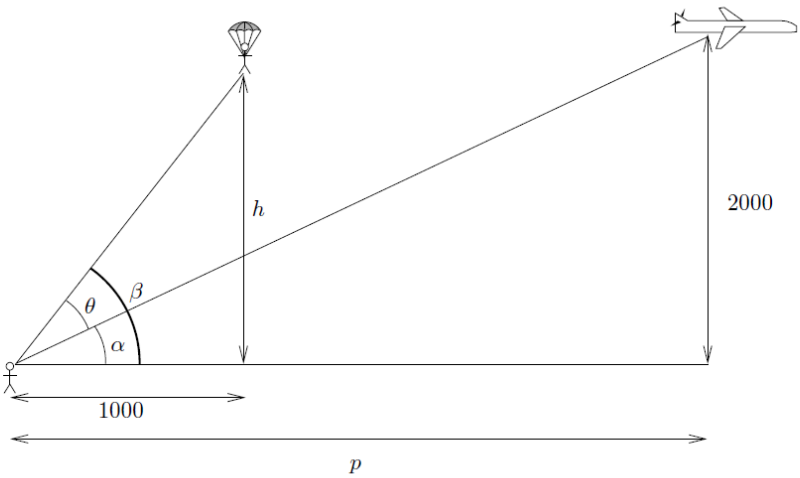
$$\frac{dv}{dt} = 8\pi - v$$

● A really hard example

Suppose that a plane is flying eastward directly away from you at a height of 2000 feet above your head. The plane moves at a constant speed of 500 feet per second. Meanwhile, some time ago a parachutist jumped out of a helicopter (which has since own away). The parachutist is floating directly downward, 1000 feet due east of you, at a constant speed of 10 feet per second. The situation is summarized in the following picture:



In the picture, what you might call the inter-azimuthal angle between the parachutist and the plane (with respect to you) is marked as θ. The question is, at what rate is θ changing when the plane and the parachutist have the same height but the plane is 8000 feet due east of you? Let the plane be p feet to the east of you. Let the height be h feet. By drawing a few extra lines, we can recast the above diagram as follows:



So we know that $\theta = \beta - \alpha$. Actually, we should probably write $\theta = |\beta - \alpha|$, just in case the parachutist is much lower than the plane. At around the time we're interested in, the heights are the same but the plane is much farther to the east than the parachutist, so β must be bigger than α and we don't need the absolute values.

Now, let's do some trig. We have two right-angled triangles

$$\tan(\alpha) = \frac{2000}{p} \qquad \text{and} \qquad \tan(\beta) = \frac{h}{1000}$$

Step 2 is finally done, and we can move on to step 3, differentiating these two relations implicitly with respect to time

$$\sec^2(\alpha)\frac{d\alpha}{dt} = -\frac{2000}{p^2}\frac{dp}{dt}$$

$$\sec^2(\beta)\frac{d\beta}{dt} = \frac{1}{1000}\frac{dh}{dt}$$

$$\theta = \beta - \alpha \qquad \frac{d\theta}{dt} = \frac{d\beta}{dt} - \frac{d\alpha}{dt}$$

Now we'd better make some substitutions and get to the bottom of this mess. Well, the speed of the plane is 500 feet per second, which means that $dp/dt = 500$. The speed of the parachutist is 10 feet per second, but the height is decreasing, so $dh/dt = -10$. We're interested in what happens when the plane is 8000 feet away, so $p = 8000$, and when the parachutist is at height 2000 feet (the same as the plane), so set $h = 2000$.
Use our trig identities, we get

$$\sec^2(\alpha) = 1 + \tan^2(\alpha) = 1 + \left(\frac{1}{4}\right)^2 = \frac{17}{16}$$

So

$$\frac{17}{16}\frac{d\alpha}{dt} = -\frac{1}{64}$$

$$\frac{d\alpha}{dt} = -\frac{1}{68}$$

The same with β

$$\sec^2(\beta) = 1 + \tan^2(\beta) = 1 + (2)^2 = 5$$

$$5\frac{d\beta}{dt} = -\frac{1}{100}$$

$$\frac{d\beta}{dt} = -\frac{1}{500}$$

So for the final equation

$$\frac{d\theta}{dt} = \frac{d\beta}{dt} - \frac{d\alpha}{dt} = \left(-\frac{1}{500}\right) - \left(-\frac{1}{68}\right) = \frac{27}{2125}$$

So the angle θ is increasing at a rate of $27/2125$ radians per second (at the moment we're considering), and we're finally done.

9.1 The Basics

● Review of exponentials

The rough idea is that we'll take a positive number, called the *base*, and raise it to a power called the *exponent*:

$$\text{base}^{\text{exponent}}$$

For example, the number $2^{-5/2}$ is an exponential with base 2 and exponent $-5/2$. It's essential that you know the so-called exponential rules, For any base $b>0$ and real numbers x and y:

1. $\boxed{b^0=1}$ The zeroth power of any nonzero number is 1

2. $\boxed{b^1=b}$ The first power of a number is just the number itself

3. $\boxed{b^x b^y=b^{x+y}}$ When you multiply two exponentials with the same base, you **add** the exponents.

4. $\boxed{\frac{b^x}{b^y}=b^{x-y}}$ When you divide two exponentials with the same base, you **subtract** the bottom exponent from the top one.

5. $\boxed{(b^x)^y=b^{xy}}$ When you take the exponential of the exponential, you **multiply** the exponents.

● Review of logarithms

Suppose that you want to solve the following equation for x:

$$2^x=7$$

The way you can bring x down from the exponent is to hit both sides with a logarithm. Since the base on the left-hand side is 2, the base of the logarithm is 2. Indeed, by definition, the solution of the above equation is

$$x=\log_2(7)$$

Let's go back to the equation $2^x=7$. We know that this means that $x=\log_2(7)$. If we now plug that value of x into the original equation, we get the bizarre looking formula

$$2^{\log_2(7)}=7$$

In more generality, $\log_b(y)$ **is the power you have to raise the base b to in order to get y**. This means that $x=\log_b(y)$ is the solution of the equation $b^x=y$ for given b and y. Plugging this value of y in, we get the formula

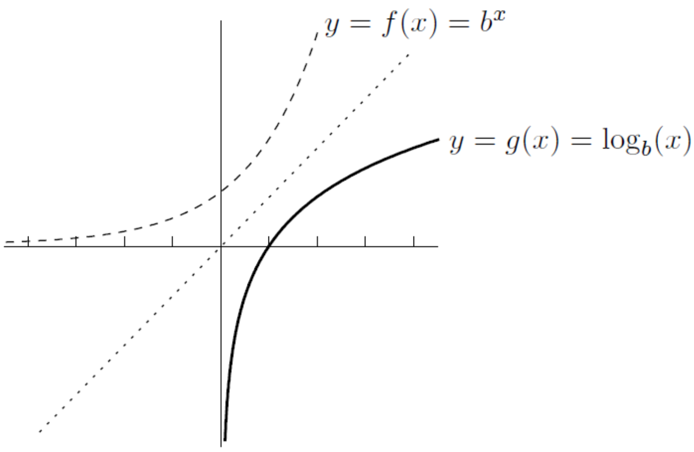
$$\boxed{b^{\log_b(y)}=y}$$

which is true for any $y>0$ and $b>0$ (except $b=1$). Remember b^x is always positive! **You can only take the logarithm of a positive number**.

You might also have noticed that I mentioned that $b=1$ is bad. For any base b between 0 and 1: $\log_b(y)=-\log_{1/b}(y)$ for all y, and $1/b$ is greater than 1.

● Logarithms, exponentials, and inverses

Fix a base $b>1$ and set $f(x)=b^x$. The function f has domain R and range $(0,\infty)$. Since it satisfies the horizontal line test, it has an inverse, which we'll call g. The domain of g is the range of f, which is $(0,\infty)$, while the range of g is the domain of f, which is $\mathbb{R}$. Remembering that the graph of the inverse function is the reflection of the original function in the mirror line $y=x$:



The exponential of the logarithm is the original number- provided that the bases match!

$$b^{\log_b(x)}=x$$

The logarithm of the exponential is the original number (provided that the bases match!)

$$\log_b(b^x)=x \quad \text{for any real } x \text{ and } b>1$$

● Log rules

Here are the rules, which are valid for any base $b>1$ and positive real numbers x and y:

1. $\boxed{\log_b(1)=0}$

2. $\boxed{\log_b(b)=1}$

3. $\boxed{\log_b(xy)=\log_b(x)+\log_b(y)}$ **The log of the product is the sum of the logs**.

4. $\boxed{\log_b(x/y)=\log_b(x)-\log_b(y)}$ **The log of the quotient is the difference of the logs**.

5. $\boxed{\log_b(x^y)=y\log_b(x)}$ **The log moves the exponent down in front of the log**. In this equation, y can be any real number (positive, negative or zero).

6. **Change of base rule**:

$$\boxed{\log_b(x)=\frac{\log_c(x)}{\log_c(b)}}$$

for any bases $b>1$ and $c>1$ and any number $x>0$. This means that all the log functions with different bases are really constant multiples of each other. Indeed, the above equation says that

$$\log_b(x)=K\log_c(x)$$

where K is constant (it happens to be equal to $1/\log_c(b)$), which means it doesn't depend on x. We can conclude that the graphs of $y=\log_b(x)$ and $y=\log_c(x)$ are very similar-you just stretch the second one vertically by a factor of K to get the first one.

Actually, there is a change of base rule for exponentials too: $b^x = c^{x \log_c(b)}$ for $b > 0$, $c > 1$, and $x > 0$.

9.2 Definition of e

So far, we haven't done any calculus involving exponentials or logs. Let's start doing some. We'll begin with limits and then move on to derivatives. Along the way, we need to introduce a new constant e, which is a special number in the same sort of way that π is a special number. One way of seeing where e comes from involves a bit of a finance lesson.

● A question about compound interest

● The answer to our question

First, let's suppose that we are compounding n times a year at an annual rate of 12%. This means that each time we compound, the amount of compounding is $0.12/n$. After this happens n times in one year, our original fortune has grown by a factor of

$$\left(1 + \frac{0.12}{n}\right)^n$$

We want to know what happens if we compound more and more often; in fact, let's allow n to get larger and larger. It would also be nice to know what happens at interest rates other than 12%. So let's replace 0.12 by r and worry about the more general limit

$$L = \lim_{n\to\infty}\left(1 + \frac{r}{n}\right)^n$$

fortune after t years, compounded n times a year at a rate of r per year $= A\left(1 + \frac{r}{n}\right)^{nt}$.

With $h = r/n$, we have

$$L = \lim_{n\to\infty}\left(1 + \frac{r}{n}\right)^n = \lim_{h\to 0^+}(1+h)^{r/h} = \lim_{h\to 0^+}\left((1+h)^{1/h}\right)^r = e^r$$

This means that if you compound more and more frequently at an annual rate of r, your fortune will increase by an amount very close to e^r, but never more than that. The quantity e^r is the "fortune-increase limit" we've been looking for. The only way you get this rate of increase is if you compound continuously-that is, all the time!

fortune after t years, compounded **continuously** at a rate of r per year $= Ae^{rt}$.

The quantities $A(1 + r/n)^{nt}$ and Ae^{rt} look quite different, but for large n they're almost the same.

● More about e and logs

Let's take a closer look at our number e. Remembering that

$$\lim_{n\to\infty}\left(1 + \frac{r}{n}\right)^n = e^r$$

we can replace r by 1 to get

$$\lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n = e$$

Of course, $r = 1$ corresponds to an interest rate of 100% per year. Let's draw up a little table of values of $(1 + 1/n)^n$ **to three decimal places** for some different values of n:

n	1	2	3	4	5	10	100	1000	10000	100000
$(1+\frac{1}{n})^n$	2	2.25	2.353	2.441	2.488	2.594	2.705	2.717	2.718	2.718

Our number e, which is the limit as $n \to \infty$ of the numbers in the second row of the above table, turns out to be an irrational number whose decimal expansion begins like this:

$$e = 2.718281828459045 23 \ldots$$

In practice, just knowing that e is a little over 2.7 will be more than enough

We can even write $\log_e(x)$ a different way: $\ln(x)$ instead of $\log_e(x)$. The expression "$\ln(x)$" is **not** pronounced "lin x" or anything like that-just say "$\log(x)$" or perhaps "ell en x", or if you're feeling particularly geeky, "the natural logarithm of x". In fact, most mathematicians write $\log(x)$ without a base to mean the same thing as $\log_e(x)$ or $\ln(x)$. The base e logarithm is called the *natural logarithm*.

Let's take another look at the log rules and formulas we've seen so far, for $x > 0$ and $y > 0$:

$$\boxed{e^{\ln(x)} = x} \quad \boxed{\ln(e^x) = x} \quad \boxed{\ln(1) = 0} \quad \boxed{\ln(e) = 1}$$

$$\boxed{\ln(xy) = \ln(x) + \ln(y)} \quad \boxed{\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)} \quad \boxed{\ln(x^y) = y\ln(x)}$$

(Actually, in the second formula, x can even be negative or 0, and in the last formula, y can be negative or 0.)

One more point before we move on to differentiating logs and exponentials. Suppose you take the important limit

$$\lim_{n\to\infty}\left(1+\frac{r}{n}\right)^n = e^r$$

and this time substitute $h = 1/n$. What we've found:

$$\boxed{\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x} \quad \text{and} \quad \boxed{\lim_{h\to 0}(1+xh)^{1/h} = e^x}$$

When $x = 1$, we get two formulas for e:

$$\boxed{\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e} \quad \text{and} \quad \boxed{\lim_{h\to 0}(1+h)^{1/h} = e}$$

These are important!

9.3 Differentiation of Logs and Exponentials

One of the reasons why the logarithm base e is called the natural logarithm is that the derivative of $\log_e(x)$ is just $1/x$.

Writing $\log_e(x)$ as $\ln(x)$, we get the important formula

$$\boxed{\frac{d}{dx}\ln(x) = \frac{1}{x}}$$

$$\boxed{\frac{d}{dx}\log_b(x) = \frac{1}{x}\log_b e = \frac{1}{x\ln(b)}}$$

Since $y = b^x$, we have proved the nice formula

$$\boxed{\frac{d}{dx}(b^x) = b^x \ln(b)}$$

$$\boxed{\frac{d}{dx}(e^x) = e^x}$$

● Examples of differentiating exponentials and logs

As long as you know the basic formulas for differentiating exponentials and logs (they are the boxed equations in the previous section), then you'll be all set.

9.4 How to Solve Limit Problems Involving Exponentials or Logs

● Limits involving the definition of e

To find or construct our classic limit

$$\lim_{h \to 0} (1+h)^{1/h} = e$$

● Behavior of exponentials near 0

In fact, since $e^0 = 1$, we know that

$$\lim_{x \to 0} e^x = e^0 = 1$$

This sort of approach works well if your exponential term appears in a product or a quotient, but it fails miserably with something like this:

$$\lim_{h \to 0} \frac{e^h - 1}{h}$$

when the dummy variable is by itself on the bottom, your limit might be a derivative in disguise:

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x)$$

$$\lim_{h \to 0} \frac{e^{x+h} - e^x}{h} = e^x$$

we get the useful fact by replacing x by 0 that

$$\boxed{\lim_{h \to 0} \frac{e^h - 1}{h} = 1}$$

● Behavior of logarithms near 1

Now let's look at how logs behave near 1. It turns out that the situation is pretty similar to the case of exponentials near 0. We know that $\ln(1) = 0$ but what is

$$\lim_{h \to 0} \frac{\ln(1+h)}{h} ?$$

This is another example of a limit which is a derivative in disguise:

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x)$$

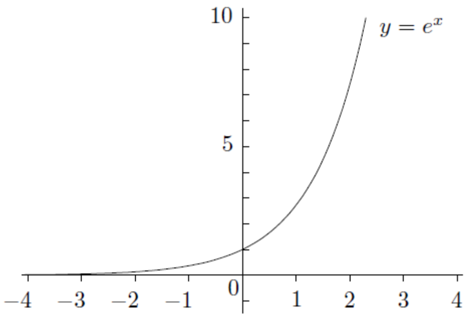
$$\lim_{h \to 0} \frac{\ln(x+h) - \ln(x)}{h} = \frac{1}{x}$$

Since $\ln(1) = 0$, this simplifies to

$$\boxed{\lim_{h \to 0} \frac{\ln(1+h)}{h} = 1}$$

● Behavior of exponentials near ∞ or $-\infty$

Now we want to understand what happens to e^x when $x \to \infty$ or $x \to -\infty$. Let's take another look at the graph of $y = e^x$:



$$\boxed{\lim_{x \to \infty} e^x = \infty \quad \text{and} \quad \lim_{x \to -\infty} e^x = 0}$$

These are special cases of the following important limit:

$$\boxed{\lim_{x \to \infty} r^x = \begin{cases} \infty & \text{if } r > 1 \\ 1 & \text{if } r = 1 \\ 0 & \text{if } 0 \leq r \leq 1 \end{cases}}$$

This is not the whole story. The limit

$$\lim_{x \to \infty} e^x = \infty$$

gets larger and larger, and is larger than

$$\lim_{x \to \infty} x^2 = \infty$$

In fact

$$\lim_{x \to \infty} \frac{x^2}{e^x} = 0$$

It is also true if you replace x^2 by any power of x. Even x^{999} can't compete with e^x
So in general we have the following principle:

Exponentials grow quickly: $\boxed{\lim_{x \to \infty} \frac{x^n}{e^x} = 0}$ no matter how large n is

In fact, by tweaking this a little, you can get a more general statement:

$$\lim_{x \to \infty} \frac{\text{poly} - \text{type stuff}}{\text{exponential of large, positive poly} - \text{type stuff}} = 0$$

For example,

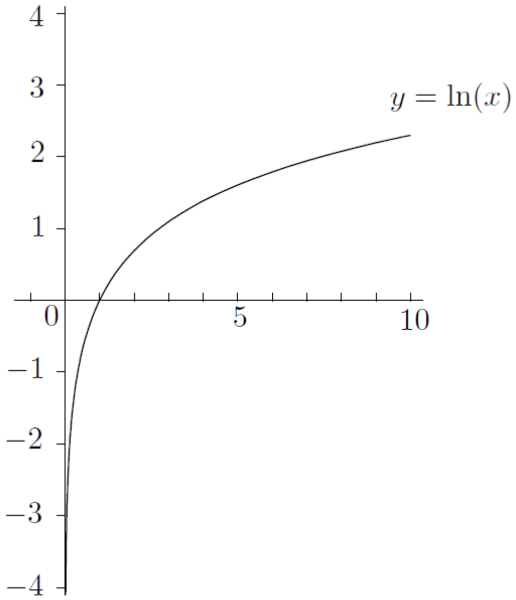
$$\lim_{x \to \infty} \frac{x^8 + 100x^7 - 4}{e^x} = 0$$

In fact, the base e can be replaced by any other base greater than 1. For example,

$$\lim_{x \to \infty} \frac{x^{10000} + 300x^9 + 32}{2^{2x^3 - 19x^2 - 100}} = 0$$

● Behavior of logs near ∞

Let's look at what happens to $\ln(x)$ when x is a large positive number. (Remember, you can't take the log of any negative number, so there's no point in studying the behavior of logs near $-\infty$) Here's the graph of $y = \ln(x)$ once again:



Again, it's important to note that the curve never touches the y-axis, even though it looks as if it does. In any event, it seems as if

$$\boxed{\lim_{x\to\infty} \ln(x) = \infty}$$

Actually, $\ln(\mathrm{x})$ goes to infinity much more slowly than any positive power of x, even something like $x^{0.0001}$. In symbols, we have

Logs grow slowly: $\boxed{\text{if } a > 0, \lim_{x\to\infty} \frac{\ln(x)}{x^a} = 0}$ no matter how small a is

Just as in the case of exponentials, it's not too hard to extend this to a more general form:

$$\lim_{x\to\infty} \frac{\text{log of positive poly} - \text{type stuff}}{\text{poly} - \text{type stuff of positive "degree"}} = 0$$

Actually, we shouldn't be surprised that logs grow slowly, once we know that exponentials grow quickly. After all, logs and exponentials are inverses of each other.

● Behavior of logs near 0

The graph of $y = \ln(x)$ above suggests that

$$\boxed{\lim_{x\to 0^+} \ln(x) = -\infty}$$

You need to use the right-hand limit here, since $\ln(x)$ isn't even defined for $x < 0$.

Consider the limit

$$\lim_{x\to 0^+} x \ln(x)$$

Here's one way to solve the above problem. Replace x by $1/t$

$$\lim_{x\to 0^+} x\ln(x) = \lim_{t\to\infty} \frac{1}{t}\ln\left(\frac{1}{t}\right) = \lim_{t\to\infty} \frac{-\ln(t)}{t} = 0$$

Logs "grow" slowly at 0: $\boxed{\text{if } a > 0,\ \lim_{x\to 0^+} x^a \ln(x) = 0}$ no matter how small a is

9.5 Logarithmic Differentiation

Logarithmic differentiation is a useful technique for dealing with derivatives of things like $f(x)^{g(x)}$, where both the base and the exponent are functions of x. After all, how on earth would you find

$$\frac{d}{dx}\left(x^{\sin(x)}\right)$$

with what we have seen already? It doesn't fit any of the rules. Still, we have these nice log rules which cut exponents down to size. If we let $y = x^{\sin(x)}$, then

$$\ln(y) = \ln\left(x^{\sin(x)}\right) = \sin(x)\ln(x)$$

by log rule. Now let's differentiate both sides (implicitly) with respect to x:

$$\frac{d}{dx}(\ln(y)) = \frac{d}{dx}(\sin(x)\ln(x))$$

Set $u = \ln(y)$, by the chain rule and product rule,

$$\frac{1}{y}\frac{dy}{dx} = \cos(x)\ln(x) + \frac{\sin(x)}{x}$$

Now we can get this:

$$\frac{dy}{dx} = \left(\cos(x)\ln(x) + \frac{\sin(x)}{x}\right)y = \left(\cos(x)\ln(x) + \frac{\sin(x)}{x}\right)x^{\sin(x)}$$

That's the answer we're looking for. (By the way, there is another way we could have done this problem. Instead of using the variable y, we could just have used our formula $A = e^{\ln(A)}$ to write

$$x^{\sin(x)} = e^{\ln\left(x^{\sin(x)}\right)} = e^{\sin(x)\ln(x)}$$

This is also work. When you've finished, you should replace $e^{\sin(x)\ln(x)}$ by $x^{\sin(x)}$ and check that you get the same answer as the original one above.

Let's review the main technique. Suppose you want to find the derivative with respect to x of

$$y = f(x)^{g(x)}$$

where both the base f and the exponent g involve the variable x. Here's what you do:

1. Let y be the function of x you want to differentiate. Take (natural) logs of both sides. The exponent g comes down on the right-hand side, so you should get

$$\ln(y) = g(x)\ln\left(f(x)\right)$$

2. Differentiate both sides implicitly with respect to x. The right-hand side often requires the product rule and the chain rule (at least). The left-hand side always works out to be $(1/y)(dy/dx)$. So you get

$$\frac{1}{y}\frac{dy}{dx} = \text{nasty stuff in } x$$

3. Multiply both sides by y to isolate dy/dx, then replace y by the original expression $f(x)^{g(x)}$, and you're done.

Even if the base and exponent are not both functions of x, logarithmic differentiation can still come in handy. If your function is really nasty and involves lots of products and quotients of powers (like x^2) and exponentials (like e^x), you might want to try logarithmic differentiation. For example,

$$\text{if } y = \frac{(x^2-3)^{100}3^{\sec(x)}}{2x^5(\log_7(x)+\cot(x))^9}, \quad \text{what is } \frac{dy}{dx}?$$

By logarithmic differentiation, that's how. Just take natural logs of both sides, and you'll find that the right-hand side becomes much more manageable.

● The derivative of x^a

Now we can finally show something that we've been taking for granted:

$$\boxed{\frac{d}{dx}(x^a) = ax^{a-1}}$$

for **any** number a, not just integers as we've seen before, when $x > 0$.

When $x \leq 0$, we have a bit of a problem. In fact, without using complex numbers, you can only make sense of x^a for $x < 0$ when a is a rational number with an odd denominator. For example, $x^{5/3}$ makes sense for negative x since you can always take a cube root-we're **OK** because 3 is odd. It's not really any different from what we've done before-just that we can handle non-integer exponents now.

9.6 Exponential Growth and Decay

We've seen that bank accounts with continuous compounding grow exponentially. It occurs in nature too. For example, under certain circumstances, populations of animals, like rabbits (and humans!), grow exponentially. There's also exponential decay, where a quantity gets smaller and smaller in an exponential fashion. This occurs in radioactive decay, allowing scientists to find out how old some ancient artifacts, fossils, or rocks are.

Here's the basic idea. Suppose $y = e^{kx}$. Then, $dy/dx = ke^{kx}$. The right-hand side of this equation can be written as ky, since $y = e^{kx}$. That is,

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

This is an example of a ***differential equation***. There are other functions satisfying the above equation. For example, if $y = 2e^{kx}$, then $dy/dx = 2ke^{kx}$, which is once again equal to ky. More generally, if $y = Ae^{kx}$, then $dy/dx = Ake^{kx}$, which is once again equal to ky. It turns out that this is the **only** way you can have $dy/dx = ky$:

$$\boxed{\text{if } \frac{dy}{dx} = ky, \quad \text{then } y = Ae^{kx} \text{ for some constant } A}$$

If we change the variable x to t, so that we are looking at $\frac{dy}{dt} = ky$ This means that the rate of change of y is equal to ky. Interesting! The rate that the quantity is changing depends on how much of the quantity you have. If you have more of the quantity, then it grows faster (assuming $k > 0$). This makes sense in the case of population growth: the more rabbits you have, the more they can breed. If you have twice as many rabbits, they also **produce** twice as many rabbits in any given time period. The number k, which is called the *growth constant*, controls how fast the rabbits are breeding in the first place. The hornier they are, the higher k

is!

● Exponential growth

So, suppose we have a population which grows exponentially. In symbols, let P (or $P(t)$, if you prefer) be the population at time t, and let k be the growth constant. The differential equation for P is

$$\frac{dP}{dt} = kP$$

We'll write P_0 to indicate that it represents the population at time 0. Altogether, we have found the

$$\boxed{\text{exponential growth equation:} \quad P(t) = P_0 e^{kt}}$$

Remember, P_0 is the initial population and k is the growth constant.

Approximation symbol: $t \cong 4\frac{2}{7}$

● Exponential decay

Let's turn things upside-down and look at exponential decay. To set the scene, let me tell you that there are certain atoms which are radioactive. They are like little time bombs: after awhile they break apart into different atoms, emitting energy at the same time. The only problem is that you never know when they are going to break apart (we'll say "decay" instead of "break apart"). All you know is that over a given time, there's a certain chance that the decay will happen.

For example, you might have a certain type of atom which has a 50% chance of decaying within any 7-year period. So if you have one of these atoms in a box, close the box, and then open it up in 7 years, there's a $50-50$ chance that it will have decayed. Of course, it's pretty difficult to see an individual atom! So let's suppose, a little more realistically, that you have a trillion atoms. (That's still a tiny speck of material, by the way.) You put them in the box and come back 7 years later. What do you expect to find? Well, about half the atoms should have decayed, while the other half remain intact. So you should have about half a trillion of the original atoms. What if you come back in another 7 years? Then half the remaining original atoms will be left, leaving you with a quarter of a trillion of the original atoms. Every 7 years, you lose half of your remaining sample.

So let's try to write down an equation to model the situation. If $P(t)$ is the number (population?) of atoms at time t, then I claim that

$$\frac{dP}{dt} = -kP$$

for some constant k. This says that the rate of change of P is a negative multiple of P

$$P(t) = P_0 e^{-kt}$$

where P_0 is the original number of atoms (at time $t = 0$). $-k$ is called the *decay constant*.

In the above example, we know that it takes 7 years for any sample of atoms to halve in size. This length of time is called the *half-life* of the atom (or material)

$$P(t) = P_0 e^{-t(\ln(2)/7)}$$

Now let's generalize a little. Suppose you have some other radioactive material with a half-life of $t_{1/2}$ years:

$$\boxed{\text{for radioactive decay with half} - \text{life } t_{1/2}, \quad P(t) = P_0 e^{-kt} \quad \text{with } k = \frac{\ln(2)}{t_{1/2}}}$$

9.7 Hyperbolic Functions

Let's change course and look at the so-called *hyperbolic functions*. These are actually exponential functions in disguise, but they are similar to trig functions in many ways. We won't be using them much but they do come up occasionally, so it's good to be familiar with them.

We'll start by defining the hyperbolic cosine and hyperbolic sine functions:

$$\boxed{\cosh(x) = \frac{e^x + e^{-x}}{2}} \qquad \boxed{\sinh(x) = \frac{e^x - e^{-x}}{2}}$$

These functions behave somewhat like their ordinary cousins, but not exactly. For example,

$$\cosh^2(x) = \frac{e^{2x} + e^{-2x} + 2}{4}$$

and

$$\sinh^2(x) = \frac{e^{2x} + e^{-2x} - 2}{4}$$

So we have

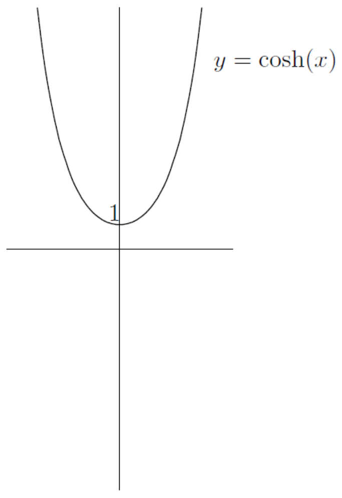
$$\boxed{\cosh^2(x) - \sinh^2(x) = 1}$$

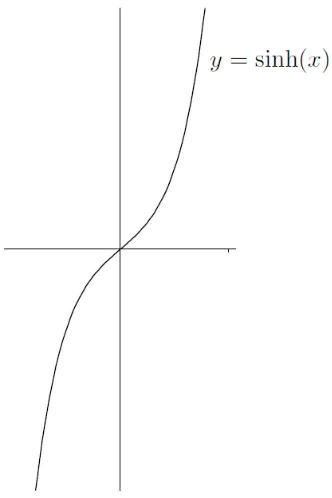
For any x. Not quite the same as the regular old trig identity-the minus makes all the difference. (Indeed, $x^2 - y^2 = 1$ is the equation of a hyperbola.)

How about calculus properties? In any case, we have

$$\boxed{\frac{d}{dx}\sinh(x) = \cosh(x)} \quad \text{and} \quad \boxed{\frac{d}{dx}\cosh(x) = \sinh(x)}$$

Now let's look at the graphs of these functions. First, you should try to convince yourself that $\cosh(x)$ is an even function of x and that $y = \sinh(x)$ is an odd function of x. (Just plug in $-x$ and see what happens.) Furthermore, $\cosh(0) = 1$ and $\sinh(0) = 0$ (check this too).





Of course you can define $\tanh(x)$ as $\sinh(x) / \cosh(x)$, as well as the reciprocals $\operatorname{sech}(x)$, $\operatorname{csch}(x)$, and $\coth(x)$.

There are also identities connecting the functions, the most important of which is

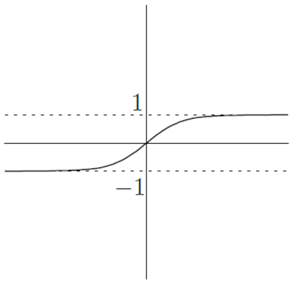
$$1 - \tanh^2(x) = \mathrm{sech}^2(x)$$

Now I'm just going to list the derivatives of the other hyperbolic functions and display their graphs

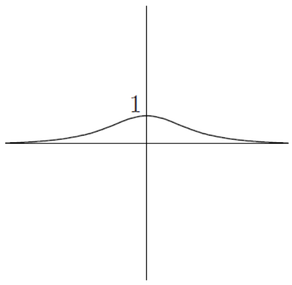
$$\boxed{\frac{d}{dx}\tanh(x) = \mathrm{sech}^2(x)} \qquad \boxed{\frac{d}{dx}\mathrm{sech}(x) = -\,\mathrm{sech}(x)\tanh(x)}$$

$$\boxed{\frac{d}{dx}\mathrm{csch}(x) = -\,\mathrm{csch}(x)\coth(x)} \qquad \boxed{\frac{d}{dx}\coth(x) = -\,\mathrm{csch}^2(x)}$$

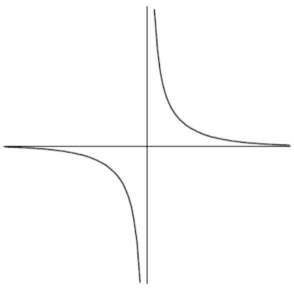
Now the graphs:



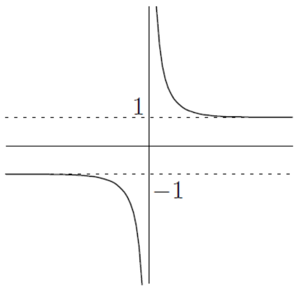
$y = \tanh(x)$



$y = \mathrm{sech}(x)$



$y = \mathrm{csch}(x)$



$y = \coth(x)$

From the definitions of the functions, you can see that all the hyperbolic trig functions are odd functions except for cosh and sech, which are even. This is the same as in the case of regular old trig functions! Also, $y = \tanh(x)$ and $y = \coth(x)$ both have horizontal asymptotes at $y = 1$ and $y = -1$, whereas $y = \mathrm{sech}(x)$ and $y = \mathrm{csch}(x)$ have a horizontal asymptote at $y = 0$

CHAPTER 10 Inverse Functions and Inverse Trig Functions

10.1 The Derivative and Inverse Functions

We're going to explore two connections between derivatives and inverse functions.

● Using the derivative to show that an inverse exists

Suppose that you have a differentiable function f whose derivative is always positive. What do you think the graph of this function looks like? Well, the slope of the tangent has to be positive everywhere, so the function can't dip up and down: it has to go upward as we look from left to right. In other words, the function must be **increasing**.

In any case, if our function f is always increasing, then it must satisfy the horizontal line test. No horizontal line could possibly hit the graph of $y = f(x)$ twice. Since the horizontal line test is satisfied by f, we know that f has an inverse. This has given us a nice strategy for showing that a function has an inverse: show that its derivative is always positive on its domain.

We've seen that if $f'(x) > 0$ for all x in the domain, then f has an inverse. There are some variations. For example, if $f'(x) < 0$ for all x, then the graph $y = f(x)$ is decreasing. The horizontal line test still works, though-the graph is just going down and down, so it can't come back up and hit the same horizontal line twice. Another variation is that the derivative might be 0 for an instant but positive everywhere else. This is **OK** as long as the derivative doesn't stay at 0 for a long time. Here's a summary of the situation:

Derivatives and inverse functions: if f is differentiable on its domain (a, b) and any of the following are true:

1. $f'(x) > 0$ for all x in (a, b);
2. $f'(x) < 0$ for all x in (a, b);
3. $f'(x) \geq 0$ for all x in (a, b) and $f'(x) = 0$ for only a finite number of x;

or

4. $f'(x) \leq 0$ for all x in (a, b) and $f'(x) = 0$ for only a finite number of x,

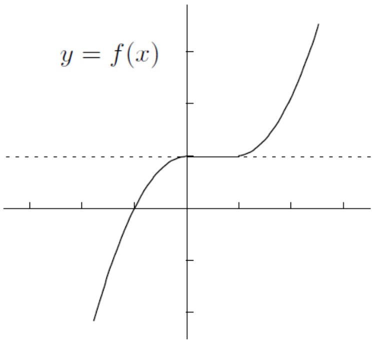
then f has an inverse. If instead the domain is of the form $[a, b]$, or $[a, b)$, or $(a, b]$, and f is continuous on the whole domain, then f still has an inverse if any of the above four conditions are true.

● Derivatives and inverse functions: what can go wrong

We noticed that the derivative of our function is allowed to be 0 occasionally and the function can still have an inverse. Why can't $f'(x) = 0$ a little more often? For example, suppose that f is defined by

$$f(x) = \begin{cases} -x^2 + 1 & \text{if } x < 0 \\ 1 & \text{if } 0 \leq x < 1 \\ x^2 - 2x + 2 & \text{if } x \geq 1 \end{cases}$$

Unfortunately the horizontal line test fails, and there is no inverse! Check out the graph



Here's another potential problem. The four conditions on the previous page all require that the domain be an interval like (a, b). What if the domain isn't in one piece? Unfortunately, then the conclusion can totally fail to hold. For example, if $f(x) = \tan(x)$, then $f'(x) = \sec^2(x)$, which can't be negative; however, you can see from the graph that $y = \tan(x)$ fails the horizontal line test pretty miserably. So the methods of the previous section won't work, in general, when your function has discontinuities or vertical asymptotes.

● Finding the derivative of an inverse function

If you know that a function f has an inverse, which we'll call f^{-1} as usual, then what's the derivative of that inverse? Here's how you find it. Start off with the equation $y = f^{-1}(x)$. You can rewrite this as $f(y) = x$. Now differentiate implicitly with respect to x to get

$$\frac{d}{dx}\big(f(y)\big) = \frac{d}{dx}(x)$$

The right-hand side is easy: it's just 1. To find the left-hand side, we use implicit differentiation.

If we set $u = f(y)$, then by the chain rule, we have $\frac{d}{dx}\big(f(y)\big) = \frac{d}{dx}(u) = \frac{du}{dy}\frac{dy}{dx} = f'(y)\frac{dy}{dx}$

Now divide both sides by $f'(y)$ to get the following principle:

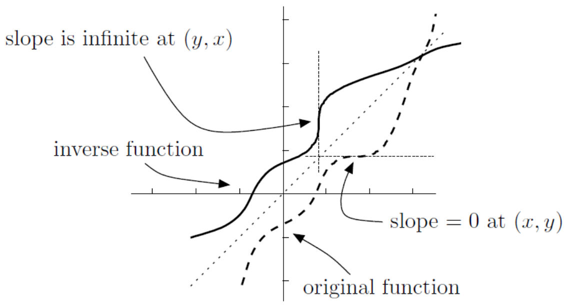
$$\boxed{\text{if } y = f^{-1}(x), \quad \text{then } \frac{dy}{dx} = \frac{1}{f'(y)}}$$

If you want to express everything in terms of x, then you have to replace y by $f^{-1}(x)$ to get

$$\boxed{\frac{d}{dx}\big(f^{-1}(x)\big) = \frac{1}{f'\big(f^{-1}(x)\big)}}$$

In words, this means that the derivative of the inverse is basically the reciprocal of the derivative of the original function.

Even though the original function is differentiable everywhere, the inverse isn't differentiable everywhere: If you have any function which has an inverse, and it has slope 0 at the point (x, y), the inverse function will have infinite slope at the point (y, x), as the following picture illustrates:

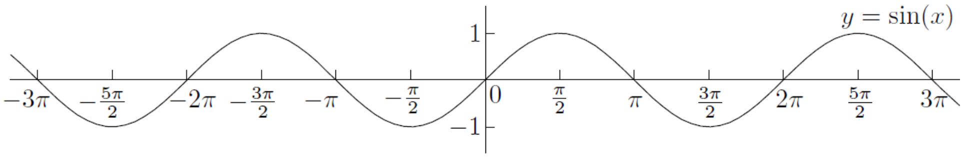


● A big example

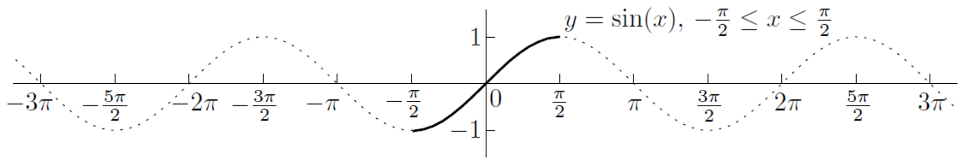
10.2 Inverse Trig Functions

● Inverse sine

Let's start by looking at the graph of $y=\sin(x)$ once again:



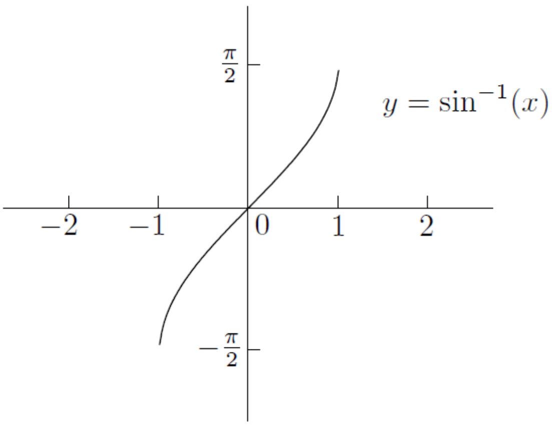
In fact, every horizontal line of height between -1 and 1 intersects the graph **infinitely** many times, which is a lot more than the zero or one time we can tolerate. So, we throw away as little of the domain as possible in order to pass the horizontal line test. There are many options, but the sensible one is to restrict the domain to the interval $[-\pi/2,\ \pi/2]$. Here's the effect of this:



OK, if $f(x)=\sin(x)$ with domain $[-\pi/2,\ \pi/2]$, then it satisfies the horizontal line test, so it has an inverse f^{-1}. We'll write $f^{-1}(x)$ as $\sin^{-1}(x)$ or $\arcsin(x)$. (Beware: the first of these notations is a little confusing at first, since $\sin^{-1}(x)$ does **not** mean the same thing as $(\sin(x))^{-1}$, even though $\sin^2(x)=(\sin(x))^2$ and $\sin^3(x)=(\sin(x))^3$.)

So, what is the domain of the inverse sine function? Well, since the range of $f(x)=\sin(x)$ is $[-1,1]$, the domain of the inverse function is $[-1,1]$. And since the domain of our function f is $[-\pi/2,\ \pi/2]$, the range of the inverse is $[-\pi/2,\ \pi/2]$.

How about the graph of $y=\sin^{-1}(x)$?



Note that since $\sin(x)$ is an odd function of x, so is $\sin^{-1}(x)$.

Now let's differentiate the inverse sine function

$$\frac{dy}{dx}=\frac{1}{\cos(y)}$$

Now, we really want the derivative in terms of x, not y. No problem-we know that $\sin(y)=x$, it shouldn't be too hard to find $\cos(y)$. In fact, $\cos^2(y)+\sin^2(y)=1$, which means that $\cos^2(y)+x^2=1$. This leads to the equation $\cos(y)=\pm\sqrt{1-x^2}$, so we have

$$\frac{dy}{dx}=\pm\frac{1}{\sqrt{1-x^2}}$$

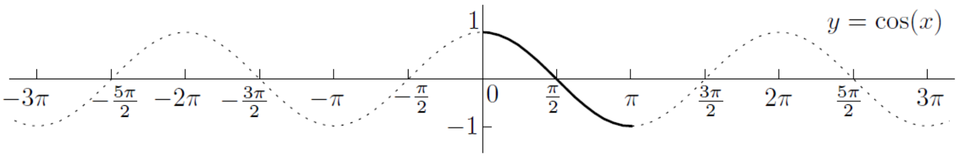
But which is it? Plus or minus? If you look at the graph of $y = \sin^{-1}(x)$ above, you can see that the slope is always positive:

$$\boxed{\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{1-x^2}} \quad \text{for} -1 < x < 1}$$

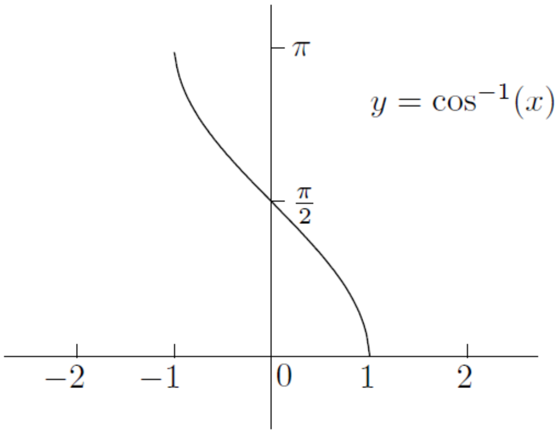
Note that $\sin^{-1}(x)$ is not differentiable, even in the one-sided sense, at the endpoints $x = 1$ and $x = -1$, since the denominator $\sqrt{1-x^2}$ is 0 in both these cases.
Here's a summary of the important facts about the inverse sine function:

$$\boxed{\sin^{-1} \text{ is odd; it has domain } [-1,1] \text{ and range } \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]}$$

● Inverse cosine
Start with the graph of $y = \cos(x)$:



This time, restricting the domain to $[-\pi/2,\ \pi/2]$ won't work, since the horizontal line test would fail and also we'd be throwing away part of the range that would be useful. You can see that the section between $[0,\ \pi]$ is highlighted and obeys the horizontal line test. We get an inverse function which we write as $\cos^{-1}$ or arccos. Like inverse sine, the domain of inverse cosine is $[-1, 1]$, since that's the range of cosine. On the other hand, the range of inverse cosine is $[0,\ \pi]$, since that's the restricted domain of cosine that we're using. The graph of $y = \cos^{-1}(x)$:



Notice that the graph shows that $\cos^{-1}$ is neither even nor odd.
Unlike the case of inverse sine, the graph of inverse cosine is all downhill, which means that the slope is always negative, so we get

$$\boxed{\frac{d}{dx}\cos^{-1}(x) = -\frac{1}{\sqrt{1-x^2}} \quad \text{for} -1 < x < 1}$$

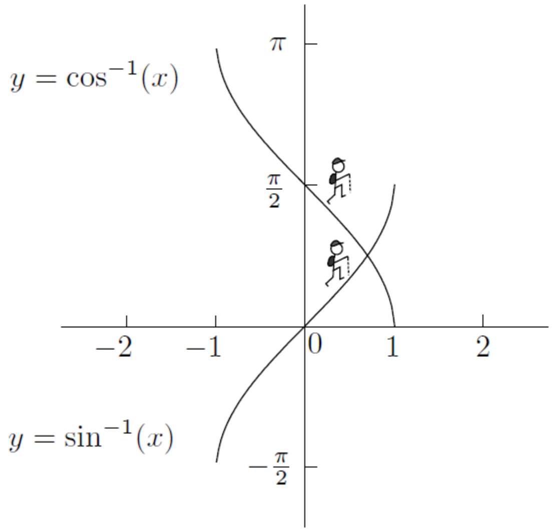
Here are the other facts about inverse cosine that we collected above:

$$\boxed{\cos^{-1} \text{ is neither even nor odd; it has domain } [-1,1] \text{ and range } [0,\pi]}$$

Let's just look at the derivatives of inverse sine and inverse cosine side by side:

$$\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{1-x^2}} \quad \text{and} \quad \frac{d}{dx}\cos^{-1}(x) = -\frac{1}{\sqrt{1-x^2}}$$

The derivatives are negatives of each other! Let's try to see why this makes sense



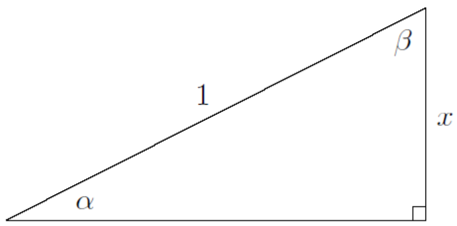
Indeed, we now know that

$$\frac{d}{dx}(\sin^{-1}(x) + \cos^{-1}(x)) = \frac{1}{\sqrt{1-x^2}} - \frac{1}{\sqrt{1-x^2}} = 0$$

So $y = \sin^{-1}(x) + \cos^{-1}(x)$ has constant slope 0, which means that it's at as a pancake. We've just used calculus to prove the following identity:

$$\sin^{-1}(x) + \cos^{-1}(x) = \frac{\pi}{2}$$

for any x in the interval $[-1, 1]$. Look at the following diagram:



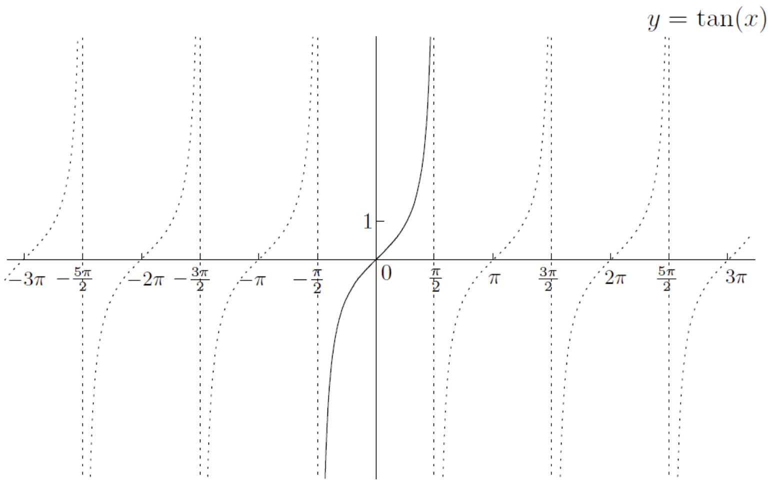
Since $\sin(\alpha) = x$, we have $\alpha = \sin^{-1}(x)$. Similarly, $\cos(\beta) = x$ which means that $\beta = \cos^{-1}(x)$. But $\alpha + \beta = \pi/2$, which means that

$$\sin^{-1}(x) + \cos^{-1}(x) = \frac{\pi}{2}$$

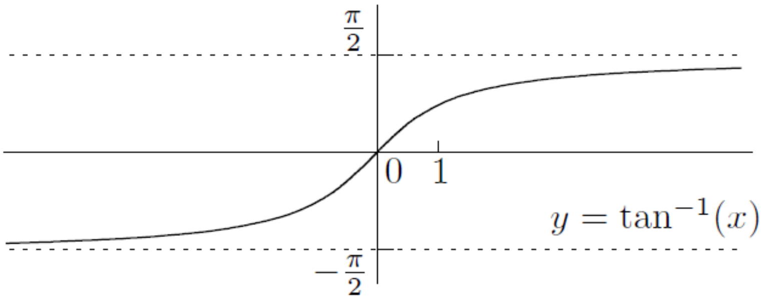
once again. Kind of nice how the calculus agrees with the geometry, huh?

● Inverse tangent

Let's remember the graph of $y = \tan(x)$:



We'll restrict the domain to $(-\pi/2,\ \pi/2)$ so that we can get an inverse function $\tan^{-1}$, also written as arctan. The domain of this function is the range of the tangent function, which is all of R. The range of the inverse function is $(-\pi/2,\ \pi/2)$. The graph of $y = \tan^{-1}(x)$ looks like this:



Now $\tan^{-1}(x)$ is an odd function of x, as you can see from the graph-it inherits its oddness from that of $\tan(x)$.

Now let's differentiate $y = \tan^{-1}(x)$ with respect to x. Write $x = \tan(y)$ and differentiate implicitly with respect to x. Since $\sec^2(y) = 1 + \tan^2(y)$, and $\tan(y) = x$, we see that $\sec^2(y) = 1 + x^2$. This means that

$$\boxed{\frac{d}{dx}\tan^{-1}(x) = \frac{1}{1+x^2} \quad \text{for all real } x}$$

We also have the following facts from above:

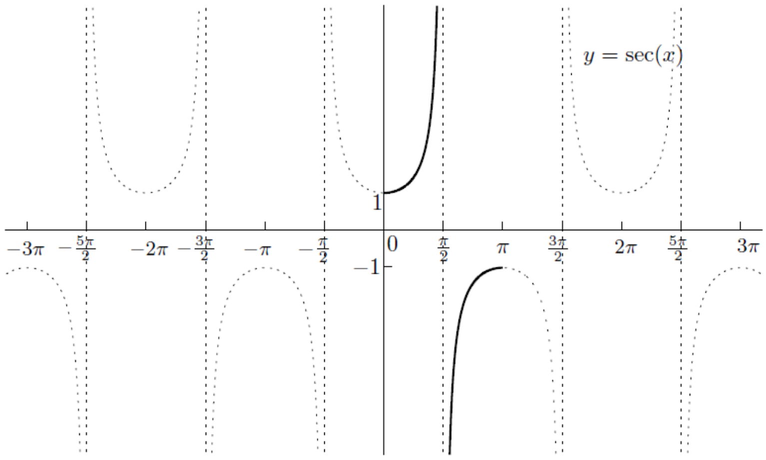
$$\boxed{\tan^{-1} \text{ is odd; it has domain R and range } \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)}$$

Unlike inverse sine and inverse cosine, the inverse tangent function has horizontal asymptotes. This means that we have the following useful limits:

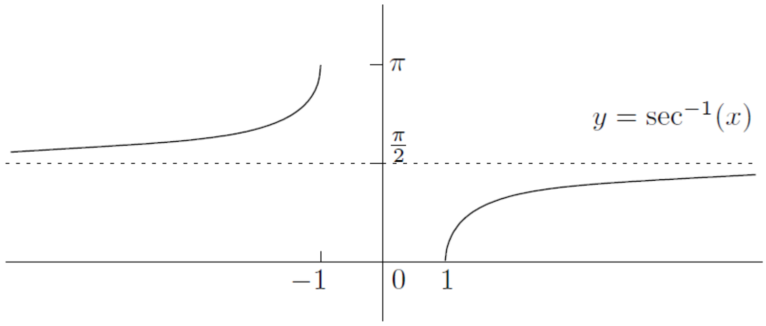
$$\boxed{\lim_{x\to\infty} \tan^{-1}(x) = \frac{\pi}{2}} \quad \text{and} \quad \boxed{\lim_{x\to-\infty} \tan^{-1}(x) = -\frac{\pi}{2}}$$

- Inverse secant

Here's the graph of $y = \sec(x)$:



The situation is (unsurprisingly) very similar to the one we faced when we inverted the cosine function. The domain has to be restricted to $[0,\pi]$, except for the point $\pi/2$, which isn't even in the original domain of $\sec(x)$. The range of secant is the union of the two intervals $(-\infty,-1]$ and $[1,\infty)$, so this becomes the domain of the inverse function $\sec^{-1}$ (alternatively arcsec). As for the range of $\sec^{-1}$, it's the same as the restricted domain: $[0,\pi]$ minus the point $\pi/2$. The graph looks like this:



Note that there's a two-sided horizontal asymptote at $y = \pi/2$, so

$$\boxed{\lim_{x\to\infty} \sec^{-1}(x) = \frac{\pi}{2}} \quad \text{and} \quad \boxed{\lim_{x\to-\infty} \sec^{-1}(x) = \frac{\pi}{2}}$$

Let's find the derivative. If $y = \sec^{-1}(x)$ then $x = \sec(y)$. So since $\sec^2(y) = 1 + \tan^2(y)$, we can rearrange and take square roots to show that $\tan(y) = \pm\sqrt{x^2-1}$. This means that

$$\frac{dy}{dx} = \frac{1}{\pm\sqrt{x^2-1}}$$

Is it plus or minus? Looking at the graph $y = \sec^{-1}(x)$ above, in fact we need to be a little more clever-instead of the plus or minus, we can simply put $|x|$ instead of x and we always get something positive. That is,

$$\boxed{\frac{d}{dx}\sec^{-1}(x) = \frac{1}{|x|\sqrt{x^2-1}} \quad \text{for } x > 1 \text{ or } x < -1}$$

We can summarize the other facts about inverse secant like this:

$$\boxed{\begin{array}{c}\sec^{-1} \text{ is neither odd nor even; it has domain} \\ (-\infty,-1] \cup [1,\infty) \text{ and range } [0,\pi]\backslash\left\{\frac{\pi}{2}\right\}\end{array}}$$

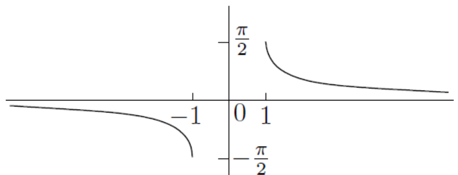
● Inverse cosecant and inverse cotangent

You can repeat the above analyses to find the domain, range, and graphs of $y = \csc^{-1}(x)$ and $y = \cot^{-1}(x)$:

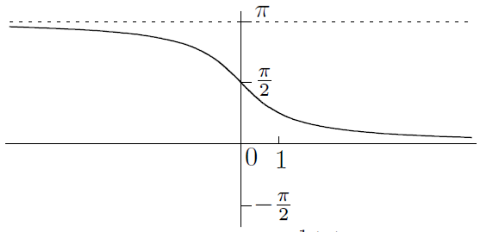
$$\boxed{\csc^{-1} \text{ is odd; it has domain } (-\infty, -1] \cup [1, \infty) \text{ and range } \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \backslash \{0\}}$$

$$\boxed{\cot^{-1} \text{ is neither odd nor even; it has domain } \mathbb{R} \text{ and range } (0, \pi)}$$

This is what the graphs look like:



$y = \csc^{-1}(x)$



$y = \cot^{-1}(x)$

Both functions have horizontal asymptotes:

$$\boxed{\lim_{x\to\infty} \csc^{-1}(x) = 0} \quad \text{and} \quad \boxed{\lim_{x\to-\infty} \csc^{-1}(x) = 0}$$

$$\boxed{\lim_{x\to\infty} \cot^{-1}(x) = 0} \quad \text{and} \quad \boxed{\lim_{x\to-\infty} \cot^{-1}(x) = \pi}$$

Notice that the graphs of $y = \csc^{-1}(x)$ and $y = \sec^{-1}(x)$ from above are very similar; in fact, you can get one from the other by flipping about the line $y = \pi/4$. This is exactly the same relation as the one that $y = \sin^{-1}(x)$ and $y = \cos^{-1}(x)$ have with each other:

$$\boxed{\frac{d}{dx}\csc^{-1}(x) = -\frac{1}{|x|\sqrt{x^2-1}} \quad \text{for } x > 1 \text{ or } x < -1}$$

$$\boxed{\frac{d}{dx}\cot^{-1}(x) = -\frac{1}{1+x^2} \quad \text{for all real } x}$$

● Computing inverse trig functions

We've completed a pretty thorough survey of the inverse trig functions. Since you have a few more derivative rules, it's a great idea to practice differentiating functions involving inverse trig functions. For one thing, you should try to make sure that you can compute quantities like $\sin^{-1}(1/2)$, $\cos^{-1}(1)$, and $\tan^{-1}(1)$ without stretching your brain:

$$\sin^{-1}(1/2) = \pi/6$$

$$\cos^{-1}(1) = 0$$

$$\tan^{-1}(1) = \pi/4$$

Now, here's some more interesting questions:

$$\sin^{-1}\left(\sin\left(\frac{13\pi}{10}\right)\right)?$$

$$\sin^{-1}\left(\sin\left(\frac{13\pi}{10}\right)\right) = -\frac{3\pi}{10}$$

$$\cos^{-1}\left(\cos\left(\frac{13\pi}{10}\right)\right)?$$

$$\cos^{-1}\left(\cos\left(\frac{13\pi}{10}\right)\right)=\frac{7\pi}{10}$$

$$\tan^{-1}\left(\tan\left(\frac{13\pi}{10}\right)\right)=\frac{3\pi}{10}$$

Just remember that tan is positive in the third quadrant!

$$\sin\left(\sin^{-1}\left(-\frac{1}{5}\right)\right)?$$

Luckily, it's not: the answer is just $-1/5$. In general, $\sin(\sin^{-1}(x))=x$, provided that x is in the domain $[-1,1]$ of inverse sine. The trouble comes when you try to write $\sin^{-1}(\sin(x))=x$. This just isn't true

$$\sin\left(\cos^{-1}\left(\frac{\sqrt{15}}{4}\right)\right)\quad\text{and}\quad\sin\left(\cos^{-1}\left(-\frac{\sqrt{15}}{4}\right)\right)?$$

The trick in both cases is to use the trig identity $\cos^2(x)+\sin^2(x)=1$

$$\sin\left(\cos^{-1}\left(\frac{\sqrt{15}}{4}\right)\right)=\frac{1}{4}$$

$$\sin\left(\cos^{-1}\left(-\frac{\sqrt{15}}{4}\right)\right)=\frac{1}{4}$$

In fact, we've noticed that $\sin(\cos^{-1}(A))$ must always be nonnegative, even if A is negative. This is because $\cos^{-1}(A)$ is in the interval $[0,\pi]$, and sine is nonnegative on that interval.

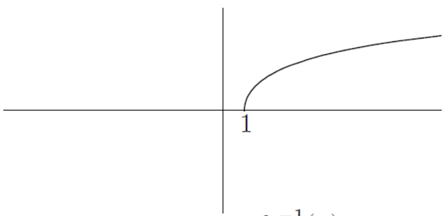
10.3 Inverse Hyperbolic Functions

The situation is a little different for hyperbolic functions. If you want an inverse for $y=\cosh(x)$, you have to throw away the left half of the graph, just as you do when you take the positive square root (and throw away the negative one). On the other hand, $y=\sinh(x)$ already satisfies the horizontal line test. So we get two inverse functions with the following properties:

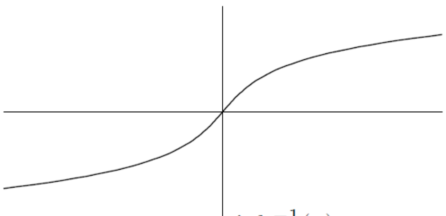
$$\boxed{\cosh^{-1}\ \text{is neither odd nor even; it has domain } [1,\infty) \text{ and range } [0,\infty)}$$

$$\boxed{\sinh^{-1}\ \text{is odd; its domain and range are all of } \mathbb{R}}$$

The graphs are obtained by reflecting the original graphs in the line $y=x$ as usual:



$y=\cosh^{-1}(x)$



$y=\sinh^{-1}(x)$

The derivatives are obtained in the same way that we got the derivatives of the inverse trig functions

$$\boxed{\frac{d}{dx}\cosh^{-1}(x)=\frac{1}{\sqrt{x^2-1}}\quad\text{for } x>1}$$

$$\boxed{\frac{d}{dx}\sinh^{-1}(x) = \frac{1}{\sqrt{x^2+1}} \quad \text{for all real } x}$$

Now, let's forget about the calculus for a few seconds and recall the definitions of $\cosh(x)$ and $\sinh(x)$:

$$\cosh(x) = \frac{e^x + e^{-x}}{2} \quad \text{and} \quad \sinh(x) = \frac{e^x - e^{-x}}{2}$$

Since we can write $\cosh(x)$ and $\sinh(x)$ in terms of exponentials, we should be able to write the inverse functions in terms of logarithms. After all, exponentials and logarithms are inverses of each other

$$\cosh^{-1}(x) = \ln\left(x + \sqrt{x^2-1}\right)$$

when $x \geq 1$

$$\sinh^{-1}(x) = \ln\left(x + \sqrt{x^2+1}\right)$$

for all x.

● The rest of the inverse hyperbolic functions

So far, we've only looked at hyperbolic sine and cosine. If you repeat the analysis for the other four hyperbolic functions, you should be able to conclude that:

$$\boxed{\tanh^{-1} \text{ is odd; its domain is } (-1,1)\text{; its range is all of } \mathrm{R}}$$

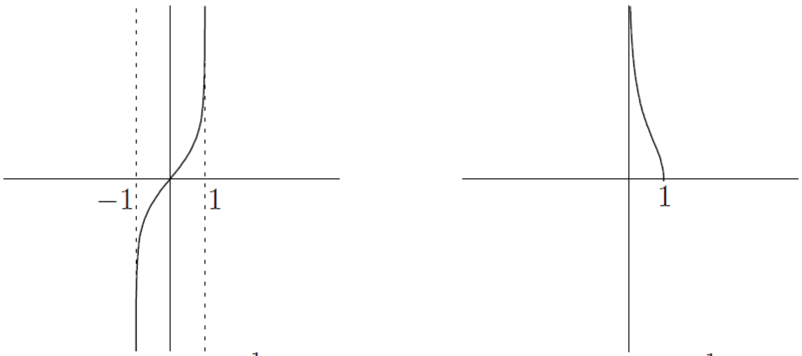
$$\boxed{\operatorname{sech}^{-1} \text{ is neither even nor odd; its domain is } (0,1]\text{; its range is } [0,\infty)}$$

$$\boxed{\operatorname{csch}^{-1} \text{ is odd; its domain and range are both } \mathrm{R}\backslash\{0\}}$$

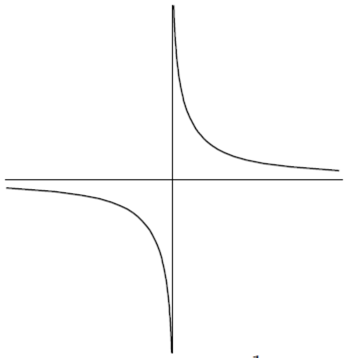
$$\boxed{\coth^{-1} \text{ is odd; its domain is } (-\infty,-1)\cup(1,\infty)\text{; its range is } \mathrm{R}\backslash\{0\}}$$

Note that we've restricted the domain of sech to $[0,\infty)$ in order to get an inverse, just as we did for cosh.

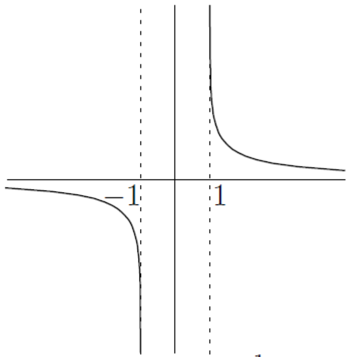
Now, here are the graphs:



$y = \tanh^{-1}(x)$ $\qquad$ $y = \operatorname{sech}^{-1}(x)$



$y = \operatorname{csch}^{-1}(x)$



$y = \coth^{-1}(x)$

Finally, you can find the derivatives using the standard trick of solving for x and differentiating implicitly with respect to x. Here's what the derivatives turn out to be:

$$\boxed{\frac{d}{dx}\tanh^{-1}(x) = \frac{1}{1-x^2} \quad (-1 < x < 1)}$$

$$\boxed{\frac{d}{dx}\coth^{-1}(x) = \frac{1}{1-x^2} \quad (x > 1 \text{ or } x < -1)}$$

$$\boxed{\frac{d}{dx}\operatorname{sech}^{-1}(x) = -\frac{1}{x\sqrt{1-x^2}} \quad (0 < x < 1)}$$

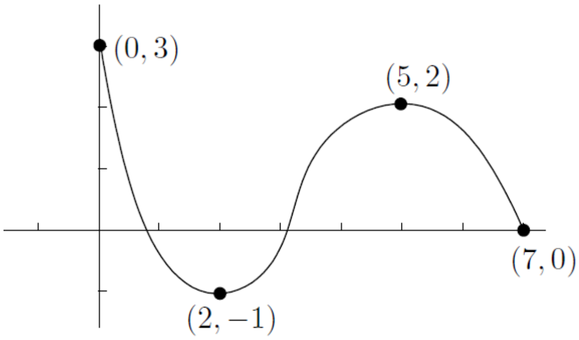
$$\boxed{\frac{d}{dx}\operatorname{csch}^{-1}(x) = -\frac{1}{|x|\sqrt{1-x^2}} \quad (x \neq 0)}$$

11.1 Extrema of Functions

If we say that x is an *extremum* of a function f, this means that f has a maximum or minimum at $x = a$. (The plural of "extremum" is "extrema," of course.) In any event, we need to go a little deeper and distinguish between two types of extrema: global and local.

Global and local extrema

The basic idea of a maximum is that it occurs when the function value is highest. Think about where the maximum of the following function on its domain $[0,7]$ should be:



Let's say that a *global maximum* (or *absolute maximum*) occurs at $x = a$ if $f(a)$ is the highest value of f on the **entire** domain of f. In symbols, we want $f(a) > f(x)$ for any value x in the domain of f. We're simply being more precise and saying "global maxima" instead of just "maxima."

As we noted before, there could be multiple global maxima; for example, $\cos(x)$ has a maximum value of 1, but this occurs for infinitely many values of x. (These values are all the integer multiples of 2π, as you can see from the graph of $y = \cos(x)$.)

How about that other type of maximum? Let's say that a *local maximum* (or *relative maximum*) occurs at $x = a$ if $f(a)$ is the highest value of f **on some small interval containing a**.

In fact, it's pretty obvious that **every global maximum is also a local maximum**.

In the same way, we can define global and local minima.

● The Extreme Value Theorem

In Chapter 5, we looked at the Max-Min Theorem. This says that a **continuous** function on a **closed** interval $[a,b]$ must have a global maximum somewhere in the interval and also a global minimum somewhere in the interval.

The problem with the Max-Min Theorem is that it doesn't tell you anything about where these global maxima and minima are. That's where the derivative comes in. Let's say that $x = c$ is a ***critical point*** for the function f if either $f'(c) = 0$ or if $f'(c)$ does not exist. Then we have this nice result:

Extreme Value Theorem: suppose that f is defined on (a,b) and c is in (a,b). If c is a **local** maximum or minimum of f, then c must be a critical point for f. That is, either $f'(c) = 0$ or $f'(c)$ does not exist.

So local maxima and minima in an open interval occur only at critical points. But it's not true that a critical point must be a local maximum or minimum! For example, $f(x) = x^3$.

The above theorem applies to open intervals. How about when the domain of your function is a closed interval $[a, b]$? Then the endpoints a and b might be local maxima and minima; they aren't covered by the theorem. So in the case of a closed interval, local maxima and minima can occur **only** at critical points or at the **endpoints** of the interval.

How to find global maxima and minima

The Extreme Value Theorem really makes finding global extrema pretty easy, since it narrows down where they can be. Here's the idea: every global extremum is also a local extremum. Local extrema can only occur at critical points. So just find all the critical points and look at the corresponding function values. The biggest one gives the global maximum, while the smallest gives the global minimum! In gory detail, here's how to find the global maximum and minimum of the function f with domain $[a, b]$:

1. Find $f'(x)$. Make a list of all the points in (a, b) where $f'(x)$ does not exist or $f'(x) = 0$. That is, make a list of all the critical points in the interval (a, b)
2. Add the endpoints $x = a$ and $x = b$ to the list
3. For each of the points in the list, find the y-coordinates by substituting into the equation $y = f(x)$
4. Pick the highest y-coordinate and note all the values of x from the list corresponding to that y-coordinate. These are the global maxima
5. Do the same for the lowest y-coordinate to find the global minima

Notice: since ∞ isn't even a number, $\lim_{x\to\infty} f(x) = A$ **can't** be an extremum.

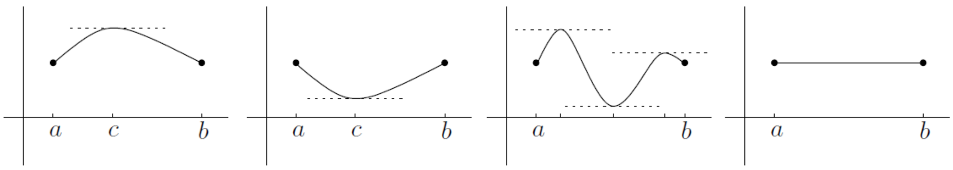
11.2 Rolle's Theorem

Imagine you're driving down a long straight highway. I watch you stop at a gas station. Then you proceed, always facing the same direction, although you can put the car in reverse if you want. Later on, I see you at the gas station again, without watching what you did in the meantime. I make the following conclusion: at some point when I wasn't looking, your car had velocity equal to zero.

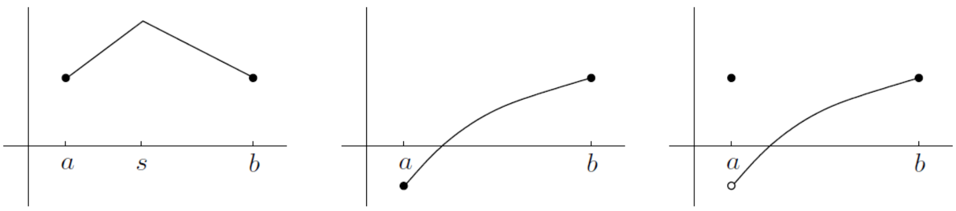
This is the content of Rolle's Theorem, which says:

Rolle's Theorem: suppose that f is continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there must be at least one number c in (a, b) such that $f'(c) = 0$.

Now, let's look at some pictures of a few possibilities of functions where Rolle's Theorem applies:



Now, let's look at some pictures where Rolle's Theorem does **not** apply:



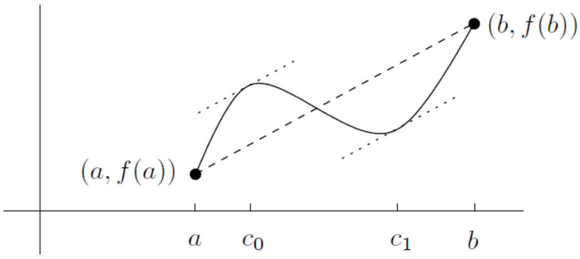
11.3 The Mean Value Theorem

Suppose you go on another journey, and I find out that you have traveled 100 miles in 2 hours. Your average velocity was 50 miles per hour. This doesn't mean that you were going at exactly 50 miles per hour the whole time. Now, here's my question: were you ever going at 50 miles per hour, even for an instant? The answer is yes. Even if you go at 45 mph for the first hour and 55 mph for the second hour, you still have to accelerate from the slow velocity to the fast velocity. This leads to the Mean Value Theorem, which says:

The Mean Value Theorem: suppose that f is continuous on $[a,b]$ and differentiable on (a,b). Then there's at least one number c in (a,b) such that

$$f'(c)=\frac{f(b)-f(a)}{b-a}$$

Let's look at a picture of the situation. Suppose your function f looks like this:



The Mean Value Theorem looks a lot like Rolle's Theorem. In fact, the conditions for applying the two theorems are almost the same. In fact, if you apply the Mean Value Theorem to a function f satisfying $f(a)=f(b)$, you'll see that $f(b)-f(a)=0$, so you get a number c in (a,b) satisfying $f'(c)=0$. So the Mean Value Theorem reduces to Rolle's Theorem!

● Consequences of the Mean Value Theorem

We've been taking a few things about the derivative for granted. For example, if a function has derivative equal to 0 everywhere, it must be constant. Facts like this seem obvious but they actually deserve to be proved. Let's use the Mean Value Theorem to show three useful facts about derivatives:

1. Suppose that a function f has derivative $f'(x)=0$ for **every** x in some interval (a,b):

if $f'(x)=0$ for all x in (a,b), then f is constant on (a,b)

2. Suppose that two differentiable functions have exactly the same derivative. Are they the same function? Not necessarily. They could differ by a constant

if $f'(x)=g'(x)$ for all x, then $f(x)=g(x)+C$ for some constant C

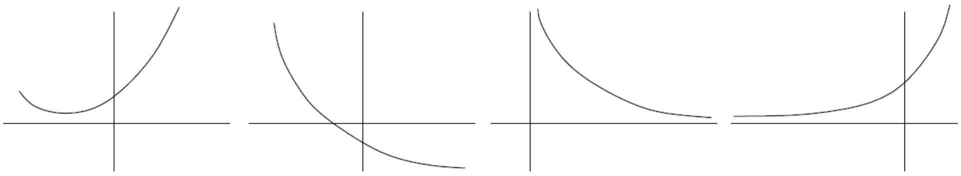
This fact will be very useful when we look at integration in a few chapters' time

3. If a function f has a derivative that's always positive, then it must be *increasing*. This means that if $a<b$, then $f(a)<f(b)$. On the other hand, if $f'(x)<0$ for all x, the function is always *decreasing*; this means that if $a<b$ then $f(b)<f(a)$.

11.4 The Second Derivative and Graphs

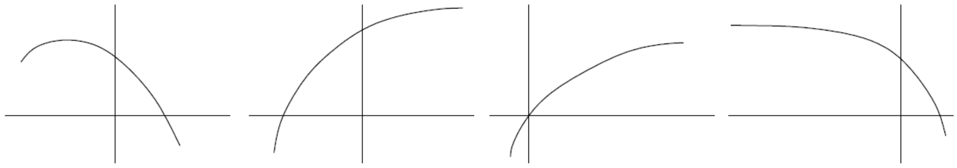
So far, we haven't paid much attention to the second derivative. We've only used it to define acceleration, and that's about all. Actually, the second derivative can tell you a lot about what the graph of your function looks like.

We'll say a function is *concave up* on an interval (a, b) if its slope is always increasing on that interval, or equivalently if its second derivative is always positive on the interval (assuming that the second derivative exists):

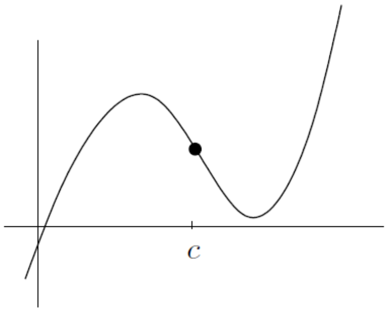


They all look like part of a bowl. Notice that you can't tell anything about the sign of the first derivative $f'(x)$ just by knowing that $f''(x) > 0$

If instead the second derivative $f''(x)$ is negative, then everything is reversed. Saying that f is *concave down* on any interval where its second derivative is always negative:



Of course, the concavity doesn't have to be the same everywhere: it can change:



We'll say that the point $x = c$ is a point of **inflection** for f because the concavity changes as you go from left to right through c.

● More about points of inflection

In the above picture, we see that $f''(x) < 0$ to the left of c and $f''(x) > 0$ to the right of c. What about $f''(x)$ itself? It must be 0, since everything is nice and smooth.

Assuming of course that $f''(x)$ actually exists when x is near c, it must be true that

if $x = c$, is a point of inflection for f, then $f''(c) = 0$

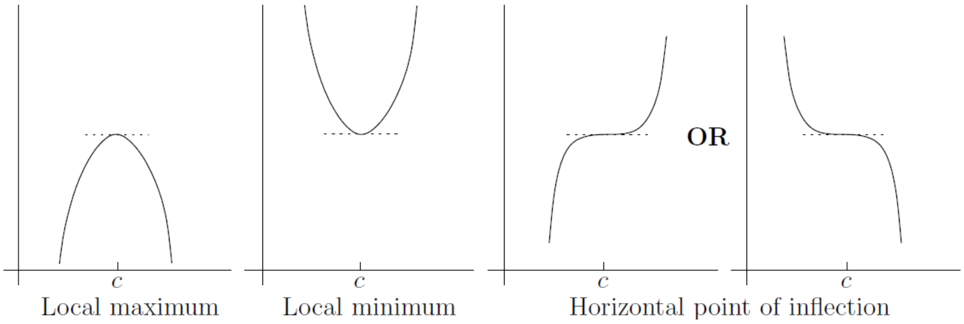
On the other hand, if $f''(c) = 0$, then c may or may not be an inflection point! That is,

if $f''(c) = 0$, then it's not always true that $x = c$ is a point of inflection for f

11.5 Classifying Points Where the Derivative Vanishes

It's time to apply some of the above theory to a practical problem. Suppose that you have a function f and a number c such that $f'(c) = 0$. You can say for sure that c is a critical point for f, but what else can you say? It turns out that there are only three common possibilities:

$x = c$ could be a local maximum; it could be a local minimum; or it could be a horizontal point of inflection, which means that it is a point of inflection with a horizontal tangent line (Another possibility is that the concavity isn't even well-defined near the critical point. For example, $f(x) = x^4 \sin(1/x)$):

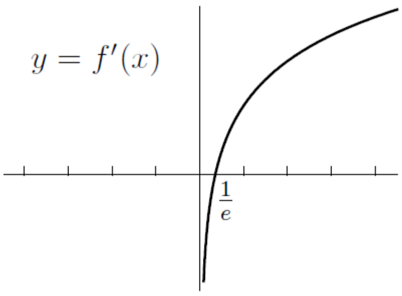


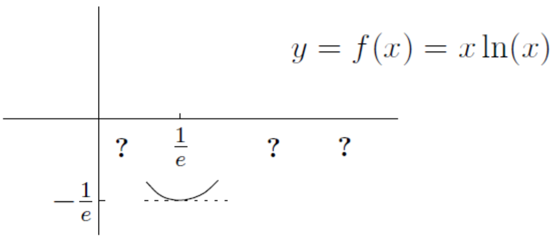
● Using the first derivative

Here's a summary of what we have just observed. Suppose that $f'(c) = 0$. Then:

1. if $f'(x)$ changes sign from positive to negative as you pass from left to right through $x = c$, then $x = c$ is a local maximum;
2. if $f'(x)$ changes sign from negative to positive as you pass from left to right through $x = c$, then $x = c$ is a local minimum;
3. if $f'(x)$ doesn't change sign as you pass through $x = c$ from left to right, then $x = c$ is a horizontal point of inflection

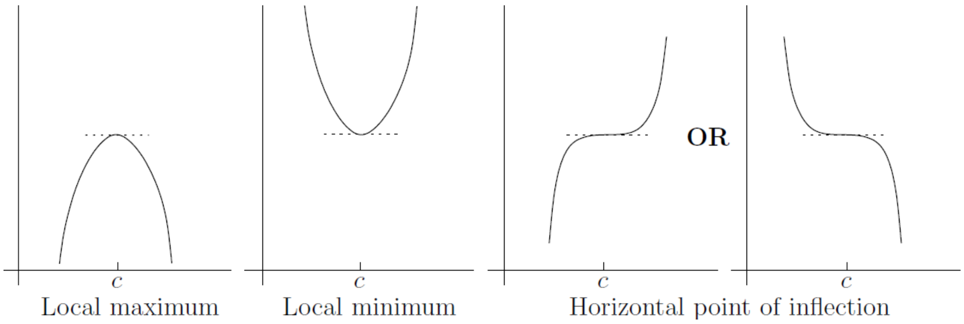
If we now set $f(x) = x\ln(x)$, then $f'(x) = \ln(x) + 1$:





● Using the second derivative

Take another look at the common possibilities which arise when $f'(c) = 0$:



Here's the summary of the situation. Suppose that $f'(c) = 0$. Then:

1. if $f''(c) < 0$, then $x = c$ is a local maximum;
2. if $f''(c) > 0$, then $x = c$ is a local minimum;
3. if $f''(c) = 0$, then you can't tell what happens! Use the first derivative test from the previous section

Yes, the first derivative test is better, although it's a little more cumbersome to use. It always works, while the second derivative test sometimes lets you down.

CHAPTER 12 Sketching Graphs

Now it's time to look at a general method for sketching the graph of $y = f(x)$ for some given function f. When we sketch a graph, we're not looking for perfection; we just want to illustrate the main features of the graph. Indeed, we're going to use the calculus tools we've developed: limits to understand the asymptotes, the first derivative to understand maxima and minima, and the second derivative to investigate the concavity.

12.1 How to Construct a Table of Signs

Suppose you want to sketch the graph of $y = f(x)$. For any number x, the quantity $f(x)$ could be positive, negative, zero, or undefined. Luckily, if f is continuous except for maybe a few points, and you can find all of the zeroes and discontinuities of f, then it's easy to see where $f(x)$ is positive and where it's negative by using **a table of signs**.

Here's how it works: start off by making a list of all the zeroes and discontinuities of f in ascending order. For example, if

$$f(x) = \frac{(x-3)(x-1)^2}{x^3(x+2)}$$

The table would look like this (with three rows and plenty of columns):

x		-2		0		1		3	
$f(x)$									

Now you can fill in some of the second row-just put a 0 where $f(x)$ is 0 and a star where f is discontinuous

x		-2		0		1		3	
$f(x)$		$\star$		$\star$		0		0	

Next, pick your favorite number between each of the special numbers on the top, as well as one at the beginning and one at the end

x	-3	-2	-1	0	$\frac{1}{2}$	1	2	3	4
$f(x)$		$\star$		$\star$		0		0	

Now, the next thing is to find whether $f(x)$ is positive or negative for each of the values we just chose. Since we could care less about the value of $f(x)$: we only care whether it's positive or negative

x	-3	-2	-1	0	$\frac{1}{2}$	1	2	3	4
$f(x)$	$-$	$\star$	$+$	$\star$	$-$	0	$-$	0	$+$

The main point is not that $f(-3)$ is negative, but that $f(x)$ is negative for **all** $x < -2$. The number -3 is just a representative sample point for the region $(-\infty, -2)$.

For now, let's see how to make a table of signs for the derivative and the second derivative.

- Making a table of signs for the derivative

A table of signs for the derivative can summarize all this information in a compact, simple way:
whenever the derivative is positive, the function is increasing; when the derivative is negative, the function is decreasing; and when the derivative is 0, the function has a local maximum, a local minimum, or a horizontal point of inflection.
The method is the same as for the table of signs for $f(x)$ that we looked at above, except that now you apply it to $f'(x)$ instead. The only other difference is that when $f'(x)$ is zero, we'll put a little at line in the third row; when $f'(x)$ is positive, the line will slope upward; and when $f'(x)$ is negative, the line will slope downward. Let's see how it works for our previous example where $f(x) = x^2(x-5)^3$. We calculated that $f'(x) = 5x(x-5)^2(x-2)$

x	-1	0	1	2	3	5	6
$f'(x)$		0		0		0	

x	-1	0	1	2	3	5	6
$f'(x)$	$+$	0	$-$	0	$+$	0	$+$
	/	—	\	_	/	—	/

A word of warning: the lines in the third row of the table are meant only to guide you as you sketch the graph of $y = f(x)$. The graph probably doesn't look like a collection of lines tacked together! Instead, just use the information in that third row to understand where the graph is increasing, decreasing or temporarily flat.

- Making a table of signs for the second derivative

The table of signs for the second derivative tells all: when the sign is positive, the curve is concave up; when the sign is negative, the curve is concave down; and when it's 0, you may or may not get a point of inflection.
The method is the same as for the function or the derivative, except that the third row is now used to show whether the function is concave up or concave down. Put a little upward parabola-like curve whenever the sign is $(+)$, a downward version when the sign is $(-)$, and a dot when the sign is 0. If we return to our example $f(x) = x^2(x-5)^3$ from above, we find that we have $f''(x) = 10(x-5)(2x^2-8x+5)$

x		$2-\frac{1}{2}\sqrt{6}$		$2+\frac{1}{2}\sqrt{6}$		5	
$f''(x)$		0		0		0	

x	0	$2-\frac{1}{2}\sqrt{6}$	2	$2+\frac{1}{2}\sqrt{6}$	4	5	6
$f''(x)$	$-$	0	$+$	0	$-$	0	$+$
	⌒	•	⌣	•	⌒	•	⌣

As we noted in the case of the first derivative in the previous section, the pictures in the third row are meant only as a guide to sketching the graph. They show where the original function is concave up and concave down, but they won't necessarily give anything more than a rough idea of what the curve $y = f(x)$ actually looks like. That's why we're going to look at a big method for sketching curves.

12.2 The Big Method

Here is an eleven-step method for sketching the graph of $y = f(x)$

1. **Symmetry**: check whether the function is even, odd, or neither by replacing x by $-x$ and seeing whether you get back the original function or its negative (you may only need to sketch it for $x \geq 0$)
2. **y-intercept**: find the y-intercept (if it exists) by setting $x = 0$
3. **x-intercepts**: find the x-intercepts by setting $y = 0$ and solving for x. This is sometimes difficult or impossible!
4. **Domain**: find the domain of f. If it's specified in the definition of f, there's nothing to do; otherwise, the domain is assumed to be as much of the real line as possible. Remember, you have to avoid numbers which lead to 0 in the denominator, or the square root of a negative number, or the log of a negative number or 0. If inverse trig functions are involved, the situation is more complicated
5. **Vertical asymptotes**: these generally occur where the denominator is zero (if there is a denominator!). Beware: if the numerator is zero too, then you might have a removable discontinuity instead of a vertical asymptote. Also, you may have a vertical asymptote due to a log factor
6. **Sign of the function**: at this point, draw up a table of signs for $f(x)$. We already know where f is zero from #3 above, and we know where it's discontinuous from #4 and #5. The table tells you exactly where the curve is above or below the x-axis
7. **Horizontal asymptotes**: find the horizontal asymptotes by calculating

$$\lim_{x\to\infty} f(x) \quad \text{and} \quad \lim_{x\to-\infty} f(x)$$

In any case, draw dashed horizontal lines on your graph to remind you about the horizontal asymptotes, if there are any

8. **Sign of the derivative**: Find the derivative, then find all the critical points-remember, these are points where the derivative is 0 or does not exist. Use the third row of the table to tell where the function is increasing, decreasing, or flat
9. **Maxima and minima**: from the table of signs, you can find all the local maxima or minima-remember, these only occur at critical points. For each maximum or minimum x, you also need to find the value of y
10. **Sign of the second derivative**: find the second derivative, then find all the points where the second derivative is zero or does not exist. The pictures in the third row of the table indicate where the curve is concave up and where it's concave down
11. **Points of inflection**: use the table of signs for the second derivative to identify the inflection points. Remember, the second derivative at an inflection point has to be zero, and the sign of the second derivative has to be different on either side of the inflection point. For each inflection point x, you need to find the y-coordinate.

12.3 Examples (Page 252)

Since the original function is odd, its derivative is even and its second derivative is odd

CHAPTER 13 Optimization and Linearization

We're now going to look at two practical applications of calculus: optimization and linearization. Basically, optimization involves finding the best situation possible, whether that be the cheapest way to build a bridge without it falling down or something as mundane as finding the fastest driving route to a specific destination. On the other hand, linearization is a useful technique for finding approximate values of hard-to-calculate quantities. It can also be used to find approximate values of zeroes of functions; this is called Newton's method.

13.1 Optimization

To “optimize” something means to make it as good as possible. This being math, we're going for quantity over quality here. The term “optimize” just means “maximize or minimize, as appropriate.”

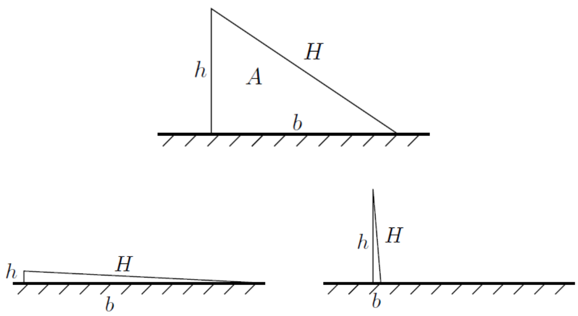
- An easy optimization example (Page 267)
- Optimization problems: the general method

Here's a way to tackle optimization problems in general:

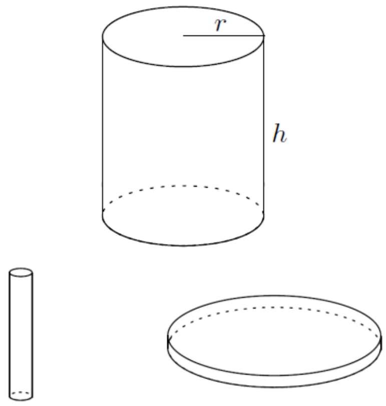
1. Identify all the variables you might possibly need. One of them should be the quantity you want to maximize or minimize-make sure you know which one! Let's call it Q for now, although of course it might be another letter like P, m, or α
2. Get a feel for the extremes of the situation, seeing how far you can push your variables
3. Write down equations relating the variables. One of them should be an equation for Q
4. Try to make Q a function of only one variable, using all your equations to eliminate the other variables
5. Differentiate Q with respect to that variable, then find the critical points; remember, these occur where the derivative is 0 or the derivative doesn't exist
6. Find the values of Q at all the critical points and at the endpoints. Pick out the maximum and minimum values. As a verification, use a table of signs or the sign of the second derivative to classify the critical points
7. Write out a summary of what you've found, identifying the variables in words rather than symbols (wherever possible)

Actually, sometimes step 4 can be quite difficult, but you might be able to avoid it altogether by using implicit differentiation

- An optimization example



- Another optimization example



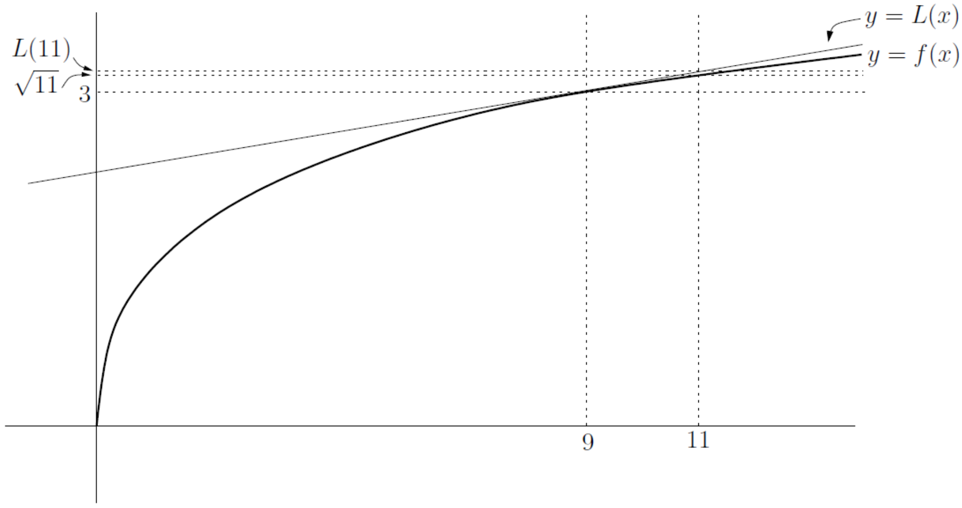
● Using implicit differentiation in optimization (Page 274)

● A difficult optimization example

13.2 Linearization

Now we're going to use the derivative to estimate certain quantities. For example, suppose you want to get a decent estimate of $\sqrt{11}$ without using a calculator. We know that $\sqrt{11}$ is a little bigger than $\sqrt{9} = 3$, so you could certainly say that $\sqrt{11}$ is approximately 3-and-a-bit. That's OK, but you can actually do a better job without too much work. Here's how it's done

Start off by setting $f(x) = \sqrt{x}$ for any $x \geq 0$. Inspired by our knowledge of $f(x)$ when $x = 9$, let's sketch the graph of $y = f(x)$, and draw in the tangent line through the point $(9, 3)$, like this:



The linear function $L(x)$ is $y = x/6 + 3/2$. This means that the value of $L(11)$ is a good approximation to $f(11) = \sqrt{11}$. We get

$$L(11) = 3\frac{1}{3}$$

We conclude that

$$\sqrt{11} \cong 3\frac{1}{3}$$

That's a lot better than 3-and-a-bit! In fact, you can use a calculator to see that $\sqrt{11}$ is 3.317 (to three decimal places), so the approximation $3\frac{1}{3}$ is pretty good.

● Linearization in general

Let's generalize the above example. If you want to estimate some quantity, try to write it as $f(x)$ for some nice function f. Next, we pick some number a, close to x (what we're interested in), such that $f(a)$ is really nice. So, given our function f and our special number a, we find the tangent to the curve $y = f(x)$ at the point $(a, f(a))$. If the tangent line is $y = L(x)$, we get

$$L(x) = f(a) + f'(a)(x - a)$$

The linear function L is called the *linearization* of f at $x = a$. Remember, we're going to use $L(x)$ as an approximation to $f(x)$. So we have

$$f(x) \cong L(x) = f(a) + f'(a)(x - a)$$

The benefit is that we now have an approximation for $f(x)$ for x **near** a.

● The differential

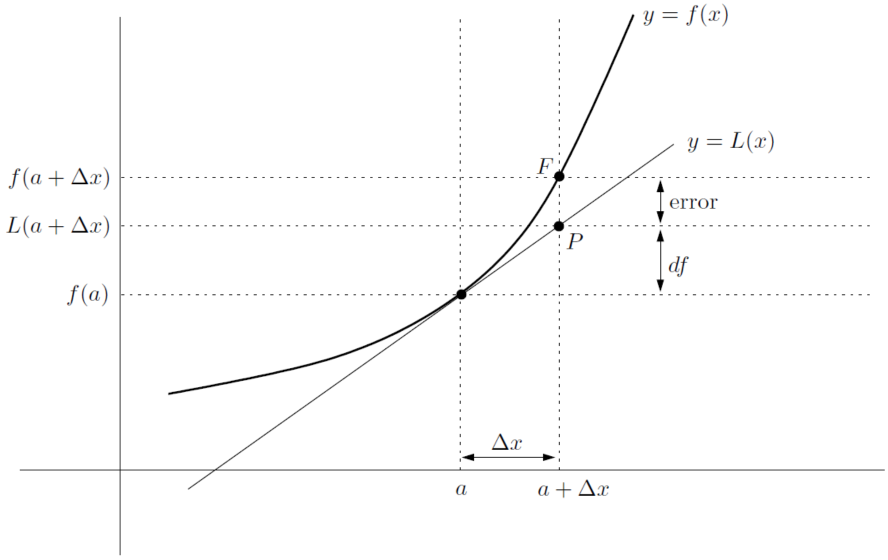
Let's take a look at the general situation once more. We saw that

$$f(x) \cong f(a) + f'(a)(x - a)$$

Let's define Δx to be $x - a$, so that $x = a + \Delta x$. The above formula becomes

$$\boxed{f(a + \Delta x) \cong f(a) + f'(a)\Delta x}$$

Here's a graph of the situation:



We want to estimate the value of $f(a + \Delta x)$. That's the height of the point F in the above picture. As an approximate value, we're actually using $L(a + \Delta x)$, which is the height of P in the picture. The difference between the two quantities is labeled "error".

In the above graph, there's one more quantity marked: this is df, which is the difference between the height of P and $f(a)$. It is the amount we needed to add to $f(a)$ in order to get our estimate. Since $L(a + \Delta x) = f(a) + f'(a)\Delta x$, we see that

$$\boxed{df = f'(a)\Delta x}$$

The quantity df is called the *differential* of f at $x = a$. It is an approximation to the amount that f changes when x moves from a to $a + \Delta x$.

We've actually touched on these ideas before: if $y = f(x)$, then $f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$

Here's an important example at page 282: a truth that when you compound the error in a one-dimensional measurement in the calculation of a three-dimensional quantity.

● Linearization summary and examples

Here's the basic strategy for estimating, or approximating, a nasty number:

1. Write down the main formula

$$\boxed{f(x) \cong L(x) = f(a) + f'(a)(x-a)}$$

2. Choose a function f, and a number x such that the nasty number is equal to $f(x)$. Also, choose a close to x such that $f(a)$ can easily be computed
3. Differentiate f to find f'
4. In the above formula, replace f and f' by the actual functions, and a by the actual number you've chosen
5. Finally, plug in the value of x from step 2 above. Also note that the differential df is the quantity $f'(a)(x-a)$

More generally an example, how would you find an approximation for $\ln(1+h)$, where h is **any** small number? That is,

$$\ln(1+h) \cong h$$

when h is small. Actually, this shouldn't be a surprise! Since $\lim_{h\to 0}\frac{\ln(1+h)}{h} = 1$

● The error in our approximation

We've been using $L(x)$ as an approximation for $f(x)$. They are not the same thing, though. How wrong could we be to use $L(x)$ instead of $f(x)$? The way to find out is to consider the difference between the two quantities:

$$r(x) = f(x) - L(x)$$

where $r(x)$ is the error in using the linearization at $x = a$ in order to estimate $f(x)$. It turns out that if the second derivative of f exists, at least between x and a, then there's a nice formula (Page 285) for $r(x)$:

$$r(x) = \frac{1}{2}f''(c)(x-a)^2 \quad \text{for some number } c \text{ between } x \text{ and } a$$

The problem is, we don't know what c is, only that it's between x and a. The above formula is related to the Mean Value Theorem.

In summary,

- if f'' is **positive** between a and x, then using the linearization leads to an **underestimate**
- if f'' is **negative** between a and x, then using the linearization leads to an **overestimate**

Now look back at the equation for the error $r(x)$ above. If we take absolute values of both sides of the equation, then we get

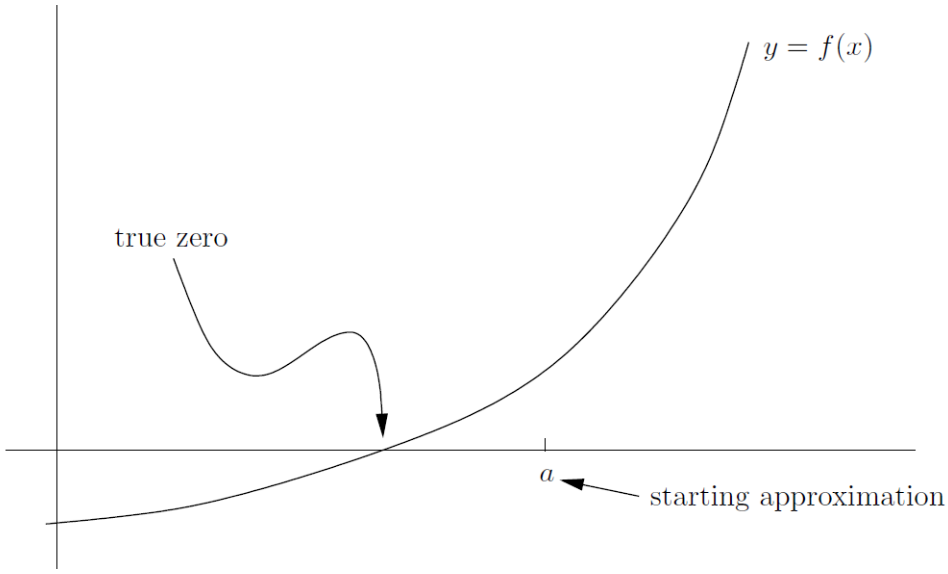
$$|\text{error}| = \frac{1}{2}|f''(c)||x-a|^2$$

Suppose we know that the biggest $|f''(c)|$ could be, as t ranges between a and x, is some number M. Then even though we don't know what c is, we do know that $|f''(c)| \leq M$, so we get the following formula:

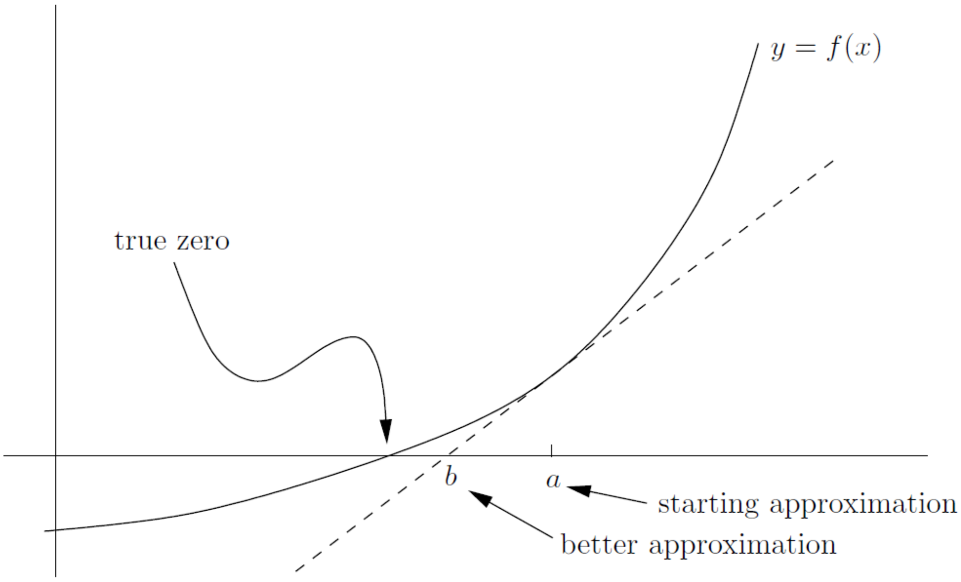
$$\boxed{|\text{error}| = \frac{1}{2}M|x-a|^2}$$

13.3 Newton's Method

Here's another useful application of linearization. Suppose that you have an equation of the form $f(x) = 0$ that you'd like to solve, but you just can't solve the darned thing. So you do the next best thing: you take a guess at a solution, which you call a. The situation might look something like this:



Think of $f(a)$ as a first stab at an approximation, which is why it's labeled "starting approximation" in the picture above. Now, the idea of Newton's method is that you can (hopefully) improve upon your estimate by using the linearization of f about $x = a$. (This means that f needs to be differentiable at $x = a$, of course!)



The x-intercept of the linearization is labeled b, and it's clearly a better approximation to the true zero than a is. So what is the value of b? Well, it's just the x-intercept of the linearization L, which is given by

$$L(x) = f(a) + f'(a)(x - a)$$

To find the x-intercept, set $L(x) = 0$; then we get

$$b = a - \frac{f(a)}{f'(a)}$$

So we have found the following formula:

> **Newton's method**: suppose that a is an approximation to a solution of $f(x) = 0$. If you set
> $$b = a - \frac{f(a)}{f'(a)}$$
> then a lot of the time b is a better approximation than a

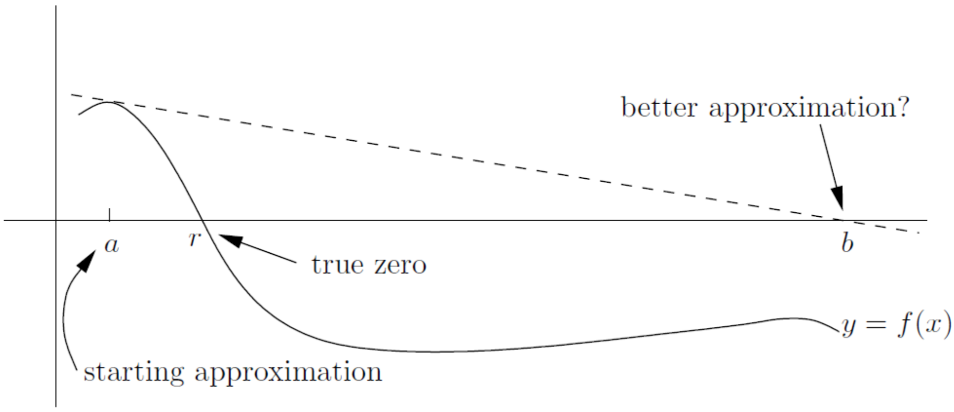
It doesn't work all the time, so I put in the phrase "a lot of the time" to cover my ass.

It might seem confusing to reuse a and b like this. A way around it is to use x_0 as the initial guess and x_1 as the first improvement; then x_2 is the second improvement, starting with x_1; and so on. The formula can now be written like this:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \quad x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}, \quad x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}, \quad \text{and so on}$$

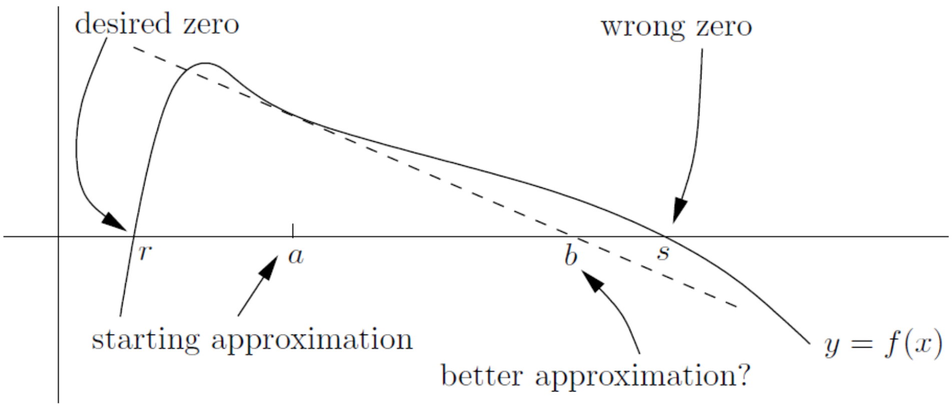
Sometimes Newton's method doesn't work. Here are four different things that could go wrong:

1. **The** value **of $f'(a)$ could be near 0**. Clearly, if
$$b = a - \frac{f(a)}{f'(a)}$$
then $f'(a)$ can't be 0 or else b isn't even defined. Even if $f'(a)$ is close, but not equal to 0, Newton's method can still give a whacked-out result; for example, check out this picture:



To get around this, make sure that your initial approximation is not near a critical point of your function f

2. **If $f(x) = 0$ has more than one solution, you might not get the right one**. For example, in the following picture, if you are trying to estimate the left-hand root r, and you guess to start at a, you'll end up estimating s instead:



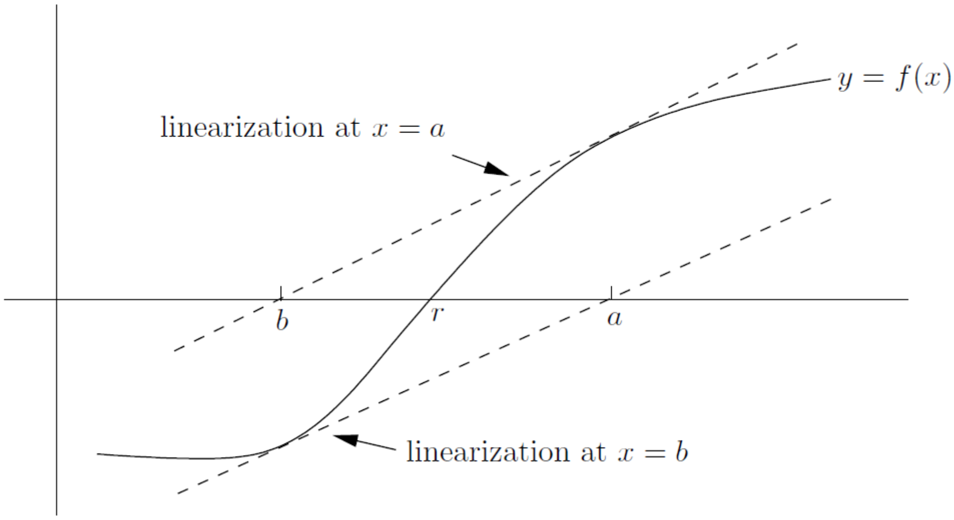
So you should make some effort to start with an estimate a which is close to the zero you want, unless you're sure there's only one zero!

3. **The approximations might get worse and worse**. For example, if $f(x) = x^{1/3}$, the only solution to the equation $f(x) = 0$ is $x = 0$. If you try to use Newton's method (for reasons best known to yourself, I guess!), then something weird happens. You see, unless you start with $a = 0$, this is what you find:

$$b = a - \frac{f(a)}{f'(a)} = -2a$$

So the next approximation is always -2 times the one you started with. These would be just getting farther and farther away from the correct value 0. There's not much you can do with Newton's method if this sort of thing happens

4. **You might get stuck in a loop**. It's possible that your estimate a leads to another estimate b, which then leads back to a again. Here's how the situation might look:



The linearization at $x = a$ has x-intercept b, and the linearization at $x = b$ has x-intercept a, so Newton's method just doesn't work. A concrete (but messy) example is

$$f(x) = \left(x^2 - \frac{4+3\pi}{4-\pi}\right)\tan^{-1}(x)$$

(By the way, the study of these sorts of loops leads to a nice type of fractal that you might have seen as a screensaver on someone's computer. . . .)

CHAPTER 14 L'Hôbpital's Rule and Overview of Limits

We've used limits to find derivatives. Now we'll turn things upside-down and use derivatives to find limits, by way of a nice technique called l'Hôpital's Rule.

14.1 L'Hôbpital's Rule

Most of the limits we've looked at are naturally in one of the following forms:

$$\lim_{x\to a}\frac{f(x)}{g(x)},\quad \lim_{x\to a}\big(f(x)-g(x)\big),\quad \lim_{x\to a} f(x)g(x),\quad \text{and}\quad \lim_{x\to a} f(x)^{g(x)}$$

Sometimes you can just substitute $x=a$ and evaluate the limit directly, effectively using the continuity of f and g. This method doesn't always work, though-for example, consider the limits

$$\lim_{x\to 3}\frac{x^2-9}{x-3},\quad \lim_{x\to 0}\left(\frac{1}{\sin(x)}-\frac{1}{x}\right),\quad \lim_{x\to 0^+} x\ln(x),\quad \text{and}\quad \lim_{x\to 0}(1+3\tan(x))^{1/x}$$

Luckily, all four types can often be solved using l'Hôpital's Rule.

It turns out that the first type, involving the ratio $f(x)/g(x)$, is the most suitable for applying the rule, so we'll call it "Type **A**." The next two types, involving $f(x)-g(x)$ and $f(x)g(x)$, both reduce directly to Type A, so we'll call them Type **B1** and Type **B2**, respectively. Finally, we'll say that limits involving exponentials like $f(x)^{g(x)}$ are Type **C**, since you can solve them by reducing them to Type **B2** and then back to Type **A**.

● Type **A**: $0/0$ case

Consider limits of the form

$$\lim_{x\to a}\frac{f(x)}{g(x)}$$

where f and g are nice differentiable functions. If $g(a)\neq 0$, everything's great. If $g(a)=0$ but $f(a)\neq 0$, then you're dealing with a vertical asymptote at $x=a$.

The only other possibility is that $f(a)=0$ and $g(a)=0$. That is, the fraction $f(a)/g(a)$ is the **indeterminate** form $0/0$.

One version of l'Hôpital's Rule:

$$\boxed{\text{if } f(a)=g(a)=0, \text{then}\quad \lim_{x\to a}\frac{f(x)}{g(x)}=\lim_{x\to a}\frac{f'(x)}{g'(x)}}$$

provided that the limit on the right-hand side exists. (Actually, there's another condition as well: $g'(x)$ can't be 0 when x is close to, but not equal to, a. You have to be really unlucky for this to be a problem, though!)

For example,

$$\lim_{h\to 0}\frac{\sqrt[5]{32+h}-2}{h}\overset{\text{l'H}}{=}\lim_{h\to 0}\frac{\frac{1}{5}(32+h)^{-4/5}}{1}=\frac{1}{5}(32)^{-4/5}$$

Notice how there's a little "l'H" above the equal sign to show that we're using l'Hôpital's Rule

$$\boxed{\overset{\text{l'H}}{=}}$$

● Type **A**: $\pm\infty/\pm\infty$ case

L'Hôpital's Rule also works in the case where $\lim_{x\to a} f(x)=\infty$ and $\lim_{x\to a} g(x)=\infty$. That is, when you try to put $x=a$, the top and bottom both look infinite, so you are dealing with the indeterminate form ∞/∞. For example

$$\lim_{x\to\infty} \frac{3x^2+7x}{2x^2-5}$$

Now, a gentle reminder: please, please, please check that you have an indeterminate form! The only acceptable forms for a quotient are $0/0$ or $\pm\infty/\pm\infty$.

● Type **B1** $(\infty - \infty)$

Here's a limit from the beginning of this chapter:

$$\lim_{x\to 0}\left(\frac{1}{\sin(x)} - \frac{1}{x}\right)$$

As $x \to 0^+$, both $1/\sin(x)$ and $1/x$ go to ∞. As $x \to 0^-$, both quantities go to $-\infty$. Either way, you're looking at the difference of two huge (positive or negative) quantities, so we can express the indeterminate form as $\pm(\infty - \infty)$.

Luckily, it's pretty easy to reduce this to Type A. Just take a common denominator:

$$\lim_{x\to 0}\left(\frac{1}{\sin(x)} - \frac{1}{x}\right) = \lim_{x\to 0}\frac{x-\sin(x)}{x\sin(x)}$$

Now we are in the $0/0$ case

$$\lim_{x\to 0}\left(\frac{1}{\sin(x)} - \frac{1}{x}\right) = \lim_{x\to 0}\frac{x-\sin(x)}{x\sin(x)} \stackrel{\text{l'H}}{=} \lim_{x\to 0}\frac{1-\cos(x)}{\sin(x)+x\cos(x)} = 0$$

Taking a common denominator doesn't always work. Sometimes you might not even have a denominator at all, so you have to create one out of thin air. For example, to find

$$\lim_{x\to\infty}\left(\sqrt{x+\ln(x)} - \sqrt{x}\right)$$

There's no denominator, so let's make one by multiplying and dividing by the conjugate expression:

$$\lim_{x\to\infty}\left(\sqrt{x+\ln(x)} - \sqrt{x}\right) = \lim_{x\to\infty}\left(\sqrt{x+\ln(x)} - \sqrt{x}\right) \times \frac{\sqrt{x+\ln(x)}+\sqrt{x}}{\sqrt{x+\ln(x)}+\sqrt{x}} = \lim_{x\to\infty}\frac{\ln(x)}{\sqrt{x+\ln(x)}+\sqrt{x}}$$

Now we are in the ∞/∞ case of Type **A**.

Unfortunately, it's not always possible to use l'Hôpital's Rule on type **B1** limits. In fact, the only time it can actually work is when you're able to manipulate the original expression to be a ratio of two quantities, as in the above examples.

● Type **B2** $(0 \times \pm\infty)$

Here's a limit we've looked at before:

$$\lim_{x\to 0^+} x\ln(x)$$

Let's turn the limit into Type **A** by manufacturing a denominator. The idea is to move x into a new denominator by putting it there as $1/x$

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

Now the form is $-\infty/\infty$, so we can use l'Hôpital's Rule.

● Type **C** $(1^{\pm\infty}, 0^0, \text{or } \infty^0)$

Finally, the trickiest type involves limits like

$$\lim_{x\to 0^+} x^{\sin(x)}$$

where both the base and exponent involve the dummy variable (x in this case). If you just put $x = 0$, you get 0^0, which is another indeterminate form. To find the limit, we'll use a technique very similar to logarithmic differentiation. The idea is to take the logarithm of the quantity $x^{\sin(x)}$ first, and work out its limit as $x \to 0^+$:

$$\lim_{x\to 0^+} \ln\left(x^{\sin(x)}\right) = \lim_{x\to 0^+} \sin(x)\ln(x)$$

As $x \to 0^+$, we have $\sin(x) \to 0$ and $\ln(x) \to -\infty$, so now we're dealing with a Type **B2** problem. We can put the $\sin(x)$ into a new denominator as $1/\sin(x)$, which is just $\csc(x)$, then use l'Hôpital's Rule on the resulting Type **A** problem:

$$\lim_{x\to 0^+} \sin(x)\ln(x) \overset{\text{l'H}}{=} \lim_{x\to 0^+} \frac{1/x}{-\csc(x)\cot(x)} = \lim_{x\to 0^+} -\frac{\sin(x)}{x} \times \tan(x) = -1 \times 0 = 0$$

Are we done? Not quite. We now know that

$$\lim_{x\to 0^+} \ln\left(x^{\sin(x)}\right) = 0$$

so now we just have to exponentiate both sides to see that

$$\lim_{x\to 0^+} x^{\sin(x)} = e^0 = 1$$

In fact, sometimes you don't even have to go through the Type **B2** step on your way to Type **A**. For example, to do

$$\lim_{x\to 0} (1 + 3\tan(x))^{1/x}$$

from the beginning of the chapter, first note that we are dealing with the form $1^{\pm\infty}$. So take logarithms:

$$\lim_{x\to 0} \ln\left((1 + 3\tan(x))^{1/x}\right) = \lim_{x\to 0} \frac{\ln(1 + 3\tan(x))}{x}$$

This is now of the form $0/0$, so it's already a Type **A** limit.

There is one more indeterminate form of this type, ∞^0. An example is

$$\lim_{x\to\infty} x^{-1/x}$$

The same trick still works: take logarithms and use the Type **A** methodology.

It's not really necessary to learn that the only indeterminate forms involving exponentials are $1^{(\pm\infty)}$, 0^0, and ∞^0. You see, if you have any limit involving exponentials, you can always use the above logarithmic method to convert everything to a product or quotient, then work out the new limit L. The actual limit will just be e^L. The only exceptions are that if $L = \infty$, then you have to interpret e^∞ as ∞; and if $L = -\infty$, then you need to recognize $e^{-\infty}$ as 0.

● Summary of l'Hôpital's Rule types

Here are all the techniques we've looked at:

1. Type **A**: if the limit involves a fraction, like

$$\lim_{x\to a} \frac{f(x)}{g(x)}$$

check that the form is indeterminate. It must be $0/0$ or ∞/∞. Use the rule

$$\boxed{\lim_{x\to a} \frac{f(x)}{g(x)} \overset{\text{l'H}}{=} \lim_{x\to a} \frac{f'(x)}{g'(x)}}$$

Do not use the quotient rule here! Now, solve the new limit, perhaps even using l'Hôpital's Rule again

2. Type **B1**: if the limit involves a difference, like

$$\lim_{x \to a} \big(f(x) - g(x)\big)$$

where the form is $\pm(\infty - \infty)$, try taking a common denominator or multiplying by a conjugate expression to reduce to a Type **A** form

3. Type **B2**: if the limit involves a product, like

$$\lim_{x \to a} f(x)g(x)$$

where the form is $0 \times \pm\infty$, pick the simplest of the two factors and put it on the bottom as its reciprocal. (Avoid picking a log term-keep that on the top.) You get something like

$$\lim_{x \to a} f(x)g(x) = \lim_{x \to a} \frac{g(x)}{1/f(x)}$$

This is now a Type **A** form

4. Type **C**: if the limit involves an exponential where both base and exponent involve the dummy variable, like

$$\lim_{x \to a} f(x)^{g(x)}$$

then first work out the limit of the logarithm:

$$\lim_{x \to a} \ln\big(f(x)^{g(x)}\big) = \lim_{x \to a} g(x) \ln\big(f(x)\big)$$

This should be either Type **B2** or Type **A** (or else it's not indeterminate and you can just substitute). Once you've solved it, you can rewrite the equation as something like

$$\lim_{x \to a} \ln\big(f(x)^{g(x)}\big) = L$$

then exponentiate both sides to get

$$\lim_{x \to a} f(x)^{g(x)} = e^L$$

14.2 Overview of Limits

It's time to consolidate. Here's a brief summary of all the techniques we've seen so far involving evaluating limits. The following techniques apply to limits of the form

$$\lim_{x \to a} F(x)$$

where F is a function which is at least continuous for x near a, but maybe not at $x = a$ itself. Also, a could be ∞ or $-\infty$. So, here's the summary:

- **Try substituting first**. You might be able to evaluate the limit
- If your substitution leads to b/∞ or $b/(-\infty)$, where b is some finite number, then the limit is 0
- If the substitution gives $b/0$, where $b \neq 0$, then you're dealing with a **vertical asymptote**. The left-hand and right-hand limits must be ∞ or $-\infty$, and the two-sided limit either doesn't exist or is one of ∞ and $-\infty$
- If none of the above points are relevant, and your limit is of the form $0/0$, try seeing if it is a **derivative in disguise**. If you can rewrite it in the form

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

● for some particular function and possibly a specific number x, then the limit is just $f'(x)$
● If **square roots** are involved, multiplication by a conjugate expression might help
● If **absolute values** are involved, convert them into piecewise-defined functions using the formula

$$|A| = \begin{cases} A & \text{if } A \geq 0 \\ -A & \text{if } A < 0 \end{cases}$$

● Otherwise, you can use the properties of various functions which can pop up as ingredients in your main function. Remember that "small" means "near 0," and "large" can mean large positive or negative numbers. Here's the deal for polynomials, trig functions, exponentials, and logs:

1. **Polynomials and poly-type functions**:
   - **General tip:** try factoring, then cancel common factors
   - **Large arguments:** the **largest-degree term dominates**, so divide and multiply by that term
2. **Trig and inverse trig functions**:
   - **General tip:** know the graphs of all the trig and inverse trig functions, and their values at some common arguments
   - **Small arguments:** $\sin(A)$ behaves like A when A is small, so divide and multiply by A. The same goes for $\tan(A)$, but **not** $\cos(A)$: that just behaves like 1. This technique is useful when only products and quotients are involved. It probably won't work when the trig function is added to or subtracted from some other quantity
   - **Large arguments:** for sine or cosine, use the facts that

   $$|\sin(\text{anything})| \leq 1 \quad \text{and} \quad |\cos(\text{anything})| \leq 1$$

   in conjunction with the sandwich principle. Some other useful facts are

   $$\lim_{x\to\infty} \tan^{-1}(x) = \frac{\pi}{2} \quad \text{and} \quad \lim_{x\to-\infty} \tan^{-1}(x) = -\frac{\pi}{2}$$

3. **Exponentials**:
   - **General tip:** know the graph of $y = e^x$, and learn the limits

   $$\lim_{h\to 0}(1 + hx)^{1/h} = e^x \quad \text{and} \quad \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n = e^x$$

   - **Small arguments:** since $e^0 = 1$, you can normally just isolate any factors which involve the exponential of a small number and replace them by 1 when you take the limit. The exception is when sums or differences occur; then you might want to use l'Hôpital's Rule, or perhaps the limit is actually a derivative in disguise
   - **Large arguments:** learn the important limits

   $$\lim_{x\to\infty} e^x = \infty \quad \text{and} \quad \lim_{x\to-\infty} e^x = 0$$

   Also remember that **exponentials grow quickly** as $x \to \infty$. This means that

   $$\lim_{x\to\infty} \frac{\text{poly}}{e^x} = 0$$

   The base e could instead be any number bigger than 1, and the exponent x could instead be some other polynomial with positive leading coefficient
4. **Logarithms**:
   - **General tip:** know the graph of $y = \ln(x)$ and the log rules

- **Small arguments:** a really important limit is

$$\lim_{x \to 0^+} \ln(x) = -\infty$$

Also, logs "grow" slowly down to $-\infty$ as $x \to 0^+$:

$$\lim_{x \to 0^+} x^a \ln(x) = 0$$

for any $a > 0$, no matter how small

- **Large arguments:** we have

$$\lim_{x \to \infty} \ln(x) = \infty$$

which has the informal abbreviation $\ln(\infty) = \infty$. Nevertheless **logs grow slowly**, that is, more slowly than any polynomial:

$$\lim_{x \to \infty} \frac{\ln(x)}{\text{poly}} = 0$$

for any polynomial of positive degree

- **Behavior near 1:** we have $\ln(1) = 0$. L'Hôpital's Rule can be very useful in this regard, or the limit might be a derivative in disguise

● If none of the above techniques work, consider using l'Hôpital's Rule. If you do, you'll always get a new limit to solve, which you can attack using any of the above principles or l'Hôpital's Rule once again.

All these facts and methods above are just tools to help you solve limits. They may not work on every limit you see. There's an art to knowing which tool to use, and of course, practice makes perfect.

CHAPTER 15 Introduction to Integration

So far as calculus is concerned, differentiation is only half the story. The other half concerns integration. This powerful tool enables us to find areas of curved regions, volumes of solids, and distances traveled by objects moving at variable speeds.

15.1 Sigma Notation

Consider the sum

$$\frac{1}{1}+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\frac{1}{25}+\frac{1}{36}$$

This is not just a sum of random numbers: there's a definite pattern. The terms in the sum are reciprocals of the squares from 1^2 through 6^2. Here's a more convenient way to write the sum:

$$\sum_{j=1}^{6}\frac{1}{j^2}$$

To read it out loud, say "the sum, from $j=1$ to 6, of $1/j^2$." We can tell that we're supposed to start at $j=1$ and end up at $j=6$ by the symbols below and above the big Greek letter Σ (which is a capital sigma, hence the term "sigma notation"). So we have

$$\sum_{j=1}^{6}\frac{1}{j^2}=\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}$$

Notice that we haven't actually worked out the value of the sum! All we've done is abbreviate it.

Sigma notation is also really useful when you want to vary where the sum stops (or starts). For example, consider the series

$$\sum_{j=1}^{n}\frac{1}{j^2}$$

This starts at $j=1$ and finishes at $j=n$, so we have

$$\sum_{j=1}^{n}\frac{1}{j^2}=\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{(n-2)^2}+\frac{1}{(n-1)^2}+\frac{1}{n^2}$$

it looks as if there are two variables, j and n, but in reality there is only one: it's n. You can easily see this by looking at the expanded form.

Let's look at some more examples. What is

$$\sum_{m=1}^{200}5\,?$$

Don't fall into the trap of saying that it's equal to 5. In fact, it is

$$\sum_{m=1}^{200}5=5+5+5+\cdots+5+5+5$$

where there are 200 terms in the sum. Similarly, consider the series

$$\sum_{q=100}^{1000}1=1+1+1+\cdots+1+1+1$$

How many terms of 1 are there in this sum? You might be tempted to say that there are $1000 - 100$, or 900, but actually there's one more. The answer is 901. In general, the number of integers between A and B, **including** A and B, is $B - A + 1$

$$\sum_{j=1}^{1501} \sin(2j-1) = \sin(1) + \sin(3) + \sin(5) + \cdots + \sin(2997) + \sin(2999) + \sin(3001)$$

$$\sum_{j=1}^{1501} \sin(2j) = \sin(2) + \sin(4) + \sin(6) + \cdots + \sin(2998) + \sin(3000) + \sin(3002)$$

● A nice sum

Consider the sum

$$\sum_{j=1}^{100} j$$

So

$$\sum_{j=1}^{100} j = 1 + 2 + 3 + \cdots + 98 + 99 + 100$$

Now, how about the sum

$$\sum_{j=0}^{99} (j+1)\,?$$

This is the same sum as before! What we've done is shift the index of summation j down by 1.

Now, consider this sum:

$$\sum_{j=1}^{100} (101-j)$$

That is

$$\sum_{j=1}^{100} (101-j) = 100 + 99 + 98 + \cdots + 3 + 2 + 1$$

This is the same sum as before, just written backward. There are many ways of expressing any sum in sigma notation.

In fact, this last way of writing the sum isn't just a curiosity-we can actually use it to find the value of the sum. Suppose that we let S be the sum $1 + 2 + \cdots + 99 + 100$; then we have seen that

$$S = \sum_{j=1}^{100} j \quad \text{and also} \quad S = \sum_{j=1}^{100} (101-j)$$

If you add up these two expressions, you get

$$2S = \sum_{j=1}^{100} j + \sum_{j=1}^{100} (101-j) = \sum_{j=1}^{100} \big(j + (101-j)\big) = \sum_{j=1}^{100} 101$$

There are 100 copies of the number 101, so we have $2S = 101 \times 100 = 10100$. This means that $S = 10100/2 = 5050$. We have shown that the sum of the numbers from 1 to 100 is 5050. Believe it or not, the great mathematician Gauss worked this out (using the same method) at the age of 10!

● Telescoping series

Check out the following sum:

$$\sum_{j=1}^{5}(j^2 - (j-1)^2)$$

This expands fully to

$$(1^2 - 0^2) + (2^2 - 1^2) + (3^2 - 2^2) + (4^2 - 3) + (5^2 - 4^2)$$

You can cancel a lot of the terms here. In fact, if you take a close look, you'll see that everything cancels out except $5^2 - 0^2$, so the sum is just $5^2 = 25$.

This sort of series is called a *telescoping series*. You can compact it down to a much simpler expression, just like collapsing one of those old spyglasses. In general, we have

$$\boxed{\sum_{j=a}^{b}\big(f(j) - f(j-1)\big) = f(b) - f(a-1)}$$

For example, we have

$$\sum_{j=10}^{100}\left(e^{\cos(j)} - e^{\cos(j-1)}\right) = e^{\cos(100)} - e^{\cos(10-1)}$$

Here's another example

$$\sum_{j=1}^{n}(j^2 - (j-1)^2) = n^2 - (1-1)^2 = n^2$$

On the other hand, the quantity $j^2 - (j-1)^2$ works out to be $j^2 - (j^2 - 2j + 1)$ or just $2j - 1$. So we have actually shown that

$$\sum_{j=1}^{n}(2j - 1) = n^2$$

So, we have proved that the sum of the first n odd numbers is n^2.

We can say even more, though. We can split up the sum like this:

$$\sum_{j=1}^{n}(2j) - \sum_{j=1}^{n}1 = n^2$$

We can pull out the constant 2 from the first sum and get

$$2\sum_{j=1}^{n}j - \sum_{j=1}^{n}1 = n^2$$

Stick the second sum on the right and divide by 2 to get

$$\sum_{j=1}^{n}j = \frac{1}{2}\left(n^2 + \sum_{j=1}^{n}1\right)$$

So the right-hand side is $(n^2+n)/2$, which can be written as $n(n+1)/2$. We have proved the useful formula

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$$

When $n = 100$, this formula specializes to

$$\sum_{j=1}^{100} j = \frac{100(100+1)}{2} = 5050$$

agreeing with what we saw in the previous section.
Instead of starting with squares as we did in the previous example, let's try starting with cubes:

$$\sum_{j=1}^{n} (j^3 - (j-1)^3) = n^3 - (1-1)^3 = n^3$$

So

$$\sum_{j=1}^{n} (3j^2 - 3j + 1) = n^3$$

$$3\sum_{j=1}^{n} j^2 - 3\sum_{j=1}^{n} j + \sum_{j=1}^{n} 1 = n^3$$

$$\sum_{j=1}^{n} j^2 = \frac{1}{3}\left(n^3 + \frac{3n(n+1)}{2} - n\right)$$

$$\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}$$

Now we know how to add up the first n square numbers. For example,

$$1^2 + 2^2 + 3^2 + \cdots + 99^2 + 100^2 = \frac{(100)(101)(201)}{6} = 338350$$

15.2 Displacement and Area

Let's move on from sigma notation, and spend some time investigating the following question: If you know the velocity of a car at every moment during some time interval, what is its total displacement over that time interval?

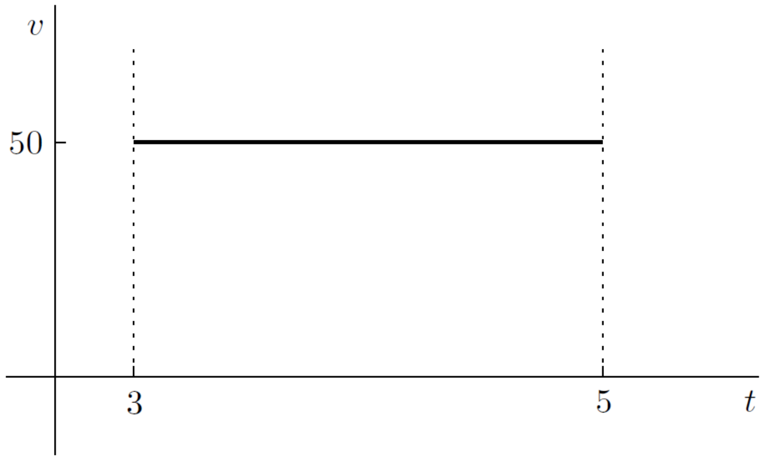
In symbols, this means that we know the velocity $v(t)$ at every time t in some interval $[a, b]$, and we want to find the displacement $x(t)$. We already know how to do this the other way around: if we know $x(t)$, then $v(t)$ is just $x'(t)$. That is, velocity is the derivative (with respect to time) of displacement. In order to answer the reverse question, let's look at some simple cases first.

● Three simple cases

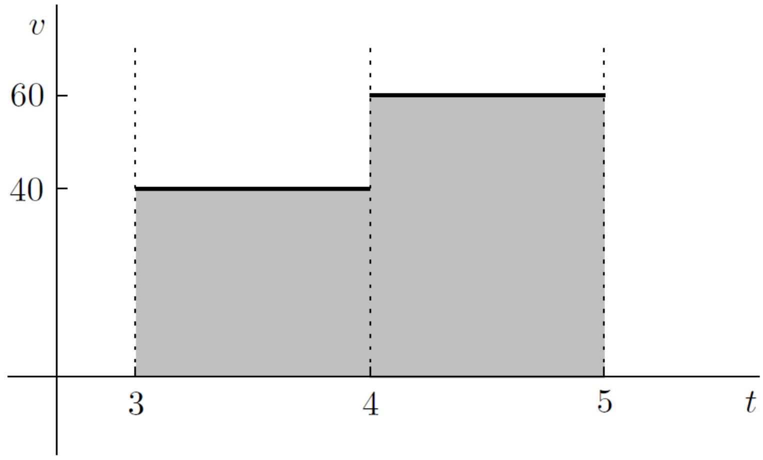
Consider three cars going in the forward direction along a long straight highway. Since the cars are always going forward, we can work with speed and distance instead of velocity and displacement (respectively)-there's no difference in this case. Each of the cars leaves from the same gas station at 3 p.m. and finishes the journey at 5 p.m

The first car goes at a speed of 50 miles per hour the whole time. So $v(t) = 50$ for all t in the interval $[3, 5]$. To work out the distance traveled in this case, just use the fact that $\text{distance} = \text{average speed} \times \text{time}$. Luckily, the average speed v_{av} and the instantaneous speed v are both equal to 50, since the speed never changes. So we get

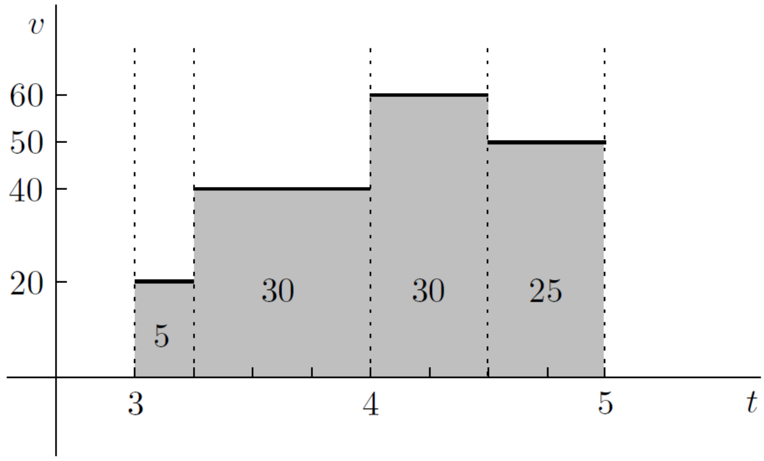
$$\text{distance} = v \times t = 50 \times 2 = 100$$



As for the second car, it goes at a speed of 40 mph for the first hour; then at 4 p.m. it starts going 60 mph. Ignoring the few seconds that it takes to accelerate, the graph of the situation looks like this:



The third car travels at 20 mph for the first 15 minutes, then goes 40 mph until 4 p.m. At that time, it switches to 60 mph for half an hour, before shifting to the slower speed of 50 mph for the rest of the journey. Once again ignoring the short accelerations and decelerations when the speed changes, the graph of v against t looks like this:



● A more general journey

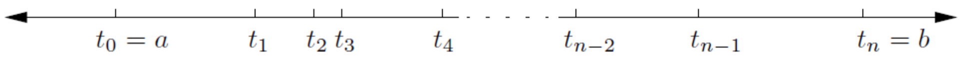
Let's look at a general framework to describe the sort of journey that the three cars made. Suppose that the time interval involved is $[a, b]$; also, suppose that we can chop up this interval into smaller intervals so that the car is going at a constant speed on each interval. We don't want to fix the number of intervals, so let's call it n. We also need to have some way of describing the beginning and end of each small interval:

1. The first interval begins at time a and finishes at some later time t_1. Since a is earlier than t_1, we can say that $a < t_1$. In fact, it will be useful to also let $t_0 = a$, so that we have $a = t_0 < t_1$
2. The second interval begins at time t_1 and finishes at some later time t_2, so that $t_1 < t_2$
3. The third interval goes from t_2 to t_3, where $t_2 < t_3$
4. Keep going in the same way, so that the jth time interval starts at time t_{j-1} and ends at time t_j
5. The second-to-last interval goes from t_{n-2} to t_{n-1}, where $t_{n-2} < t_{n-1}$
6. Finally, the last interval goes from t_{n-1} to t_n, which is the same as the very end time b. So we have $t_{n-1} < t_n = b$.

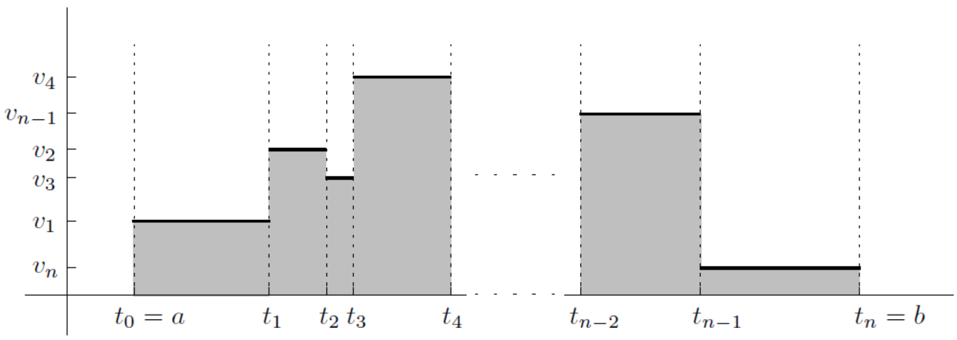
All together, we can summarize the situation by saying that

$$a = t_0 < t_1 < t_2 < t_3 < \cdots < t_{n-2} < t_{n-1} < t_n = b$$

On the number line, it looks something like this:



Overall, the picture looks like this (for example):

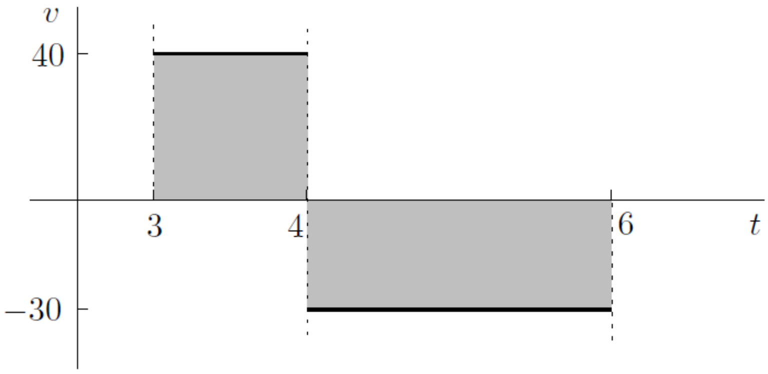


$$\text{total displacement} = v_1(t_1 - t_0) + v_2(t_2 - t_1) + \cdots + v_{n-1}(t_{n-1} - t_{n-2}) + v_n(t_n - t_{n-1})$$

$$\text{total displacement} = \sum_{j=1}^{n} v_j(t_j - t_{j-1})$$

● Signed area

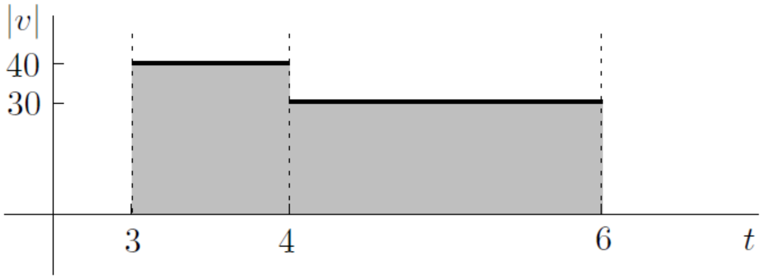
What if our car goes backward? For example, suppose that the car goes forward at 40 mph between 3 and 4 p.m., then backward at 30 mph until 6 p.m. The graph looks like this:



Now it's really important to distinguish between distance and displacement. The total distance traveled from 3 p.m. to 6 p.m. is $40 + 60 = 100$ miles. On the other hand, the displacement is $40 + (-60) = -20$ miles, since the second part of the journey is backward.

Of course, a rectangle can't actually have a negative height, but nevertheless it would be good to distinguish between rectangles above and below the axis. So if the "height" is -30 mph, then the "area" is $2 \times (-30) = -60$ miles. Let's drop the quotation marks and correctly refer to this as the *signed area*. Our convention, then, is that areas below the axis count as negative toward the total.

So instead of adding up the actual (unsigned) area in the graph above to get the distance, we could graph $|v|$ against t:

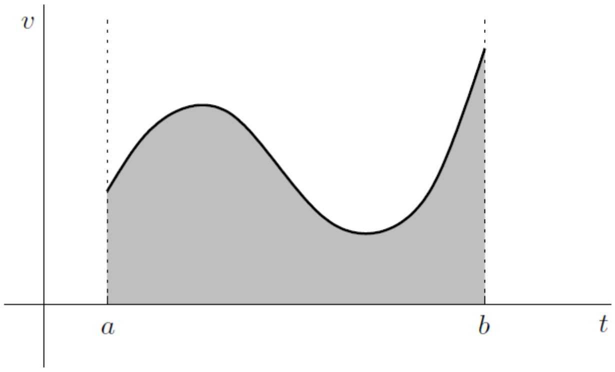


Now it's irrelevant whether the area is signed or not because there's nothing below the horizontal axis! So, we'll make the convention that **all areas are signed**. If we want the unsigned area, we'll take absolute values first.

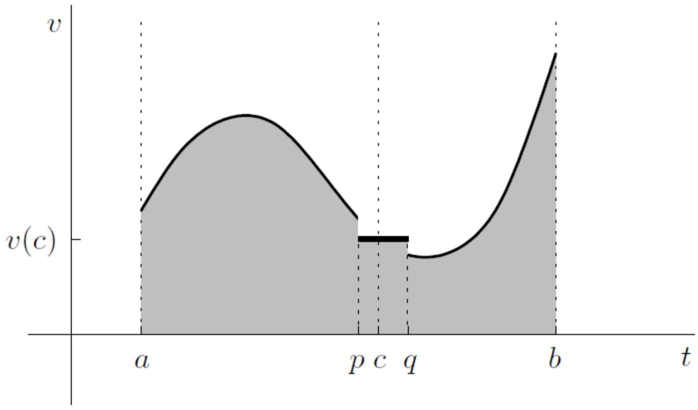
● Continuous velocity

We've seen that if a car (or other object) moves along a straight line so that the velocity is constant on a finite number of intervals in a partition of $[a, b]$, then the displacement is the signed area between the graph of v versus t, the t-axis, and the lines $t = a$ and $t = b$. The distance is the same thing, except that you start with the graph of $|v|$ versus t instead.

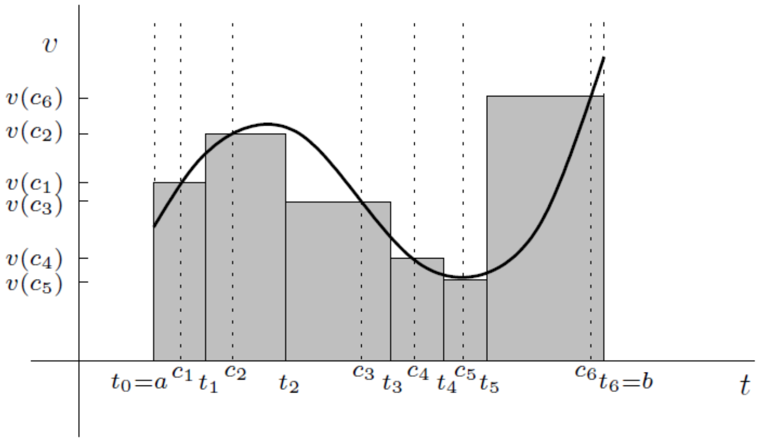
What if the velocity **isn't** constant on a finite number of intervals? Unless you never turn off the cruise control, you'll be speeding up from time to time to pass another car, slowing down when you see a cop, and so on. Even getting from 40 to 60 mph requires some acceleration-you can't just change speeds instantaneously. So, let's consider the situation where velocity v is a **continuous** function of time t, for example:



Here's the idea. Let's **sample** the velocity by picking some instant of time c during $[p,q]$



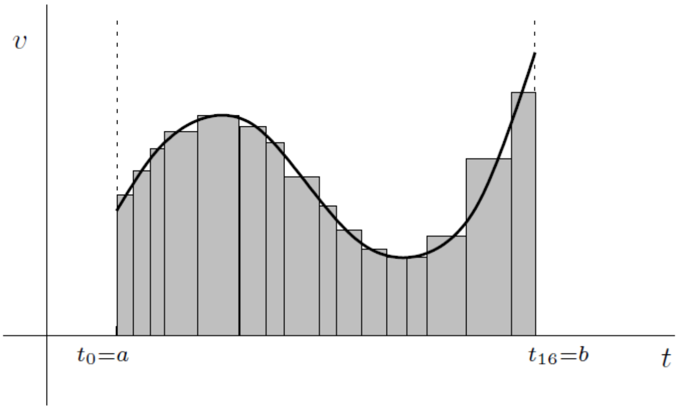
The number c_1 could be equal to the beginning number t_0 or the end number t_1, or some number in between, as long as it lies in $[t_0, t_1]$



All we've done is approximate the nice smooth velocity curve using some staircase-like function, where each step intersects the curve. We can use the techniques from the previous sections to work out the shaded (signed) area, which will be an approximation to the actual area under the curve. We get

$$\text{area under velocity curve} \cong \sum_{j=1}^{n} v(c_j)(t_j - t_{j-1})$$

Unfortunately, the approximation is pretty lousy. So let's take a different partition with more intervals which are smaller, for example:



So somehow we need to make **all** the little time intervals small. The way to do this is to let the mesh of the partition be the longest of all the time intervals, then insist that the mesh get smaller and smaller, eventually down to 0 in the limit.

Formally, the mesh is defined by

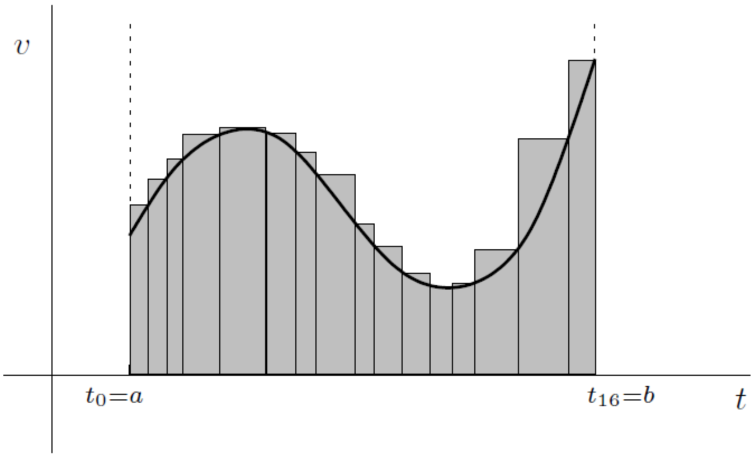
$$\text{mesh} = \text{maximum of } (t_1 - t_0), (t_2 - t_1), \cdots, (t_{n-1} - t_{n-2}), (t_n - t_{n-1})$$

This is what we're trying to achieve in the following formula:

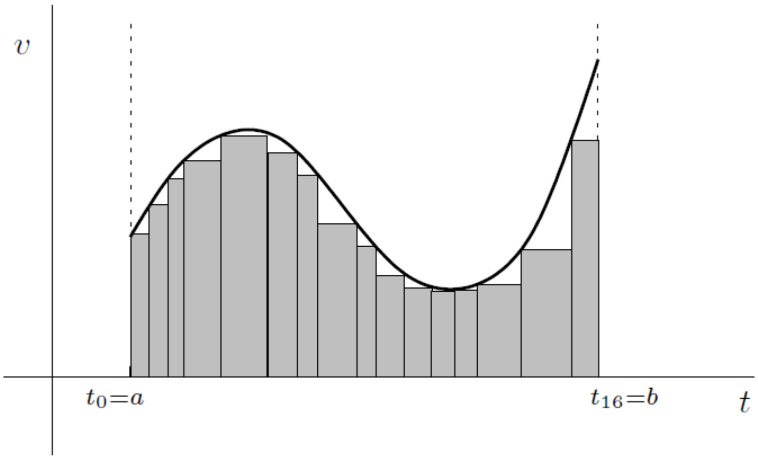
$$\text{actual area under velocity curve} = \lim_{\text{mesh}\to 0} \sum_{j=1}^{n} v(c_j)(t_j - t_{j-1})$$

● Two special approximations

The above formula leaves a lot to be desired. How do you know that you get the same answer no matter what partitions you take and no matter how you choose the sampling times c_j. It's actually a theorem that if v is a continuous function of t, then the above limit is independent of the partitions and sampling times. We can get an idea of the flavor of the proof by investigating two special approximations: the upper sum and the lower sum



The area of the rectangles, which is called *an upper sum*, is clearly bigger than the area under the curve



The area of all the rectangles, which is called *a lower sum*, is less than the area under the curve. Combining these observations, we have

$$\text{lower sum} \le \text{actual area under curve} \le \text{upper sum}$$

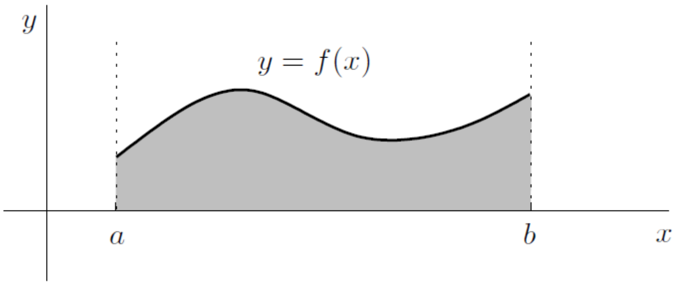
If you use a sequence of partitions with smaller and smaller meshes, then the lower sum and the upper sum have the same limit (that's what I'm not going to prove). The sandwich principle then shows that the formula at the end of the previous section makes sense.

CHAPTER 16 Definite Integrals

Now it's time to get some facts straight about **definite integrals**.

16.1 The Basic Idea

We start off with some function f and an interval $[a, b]$. Take the graph of $y = f(x)$, and consider the region between the curve, the x-axis, and the two vertical lines $x = a$ and $x = b$:

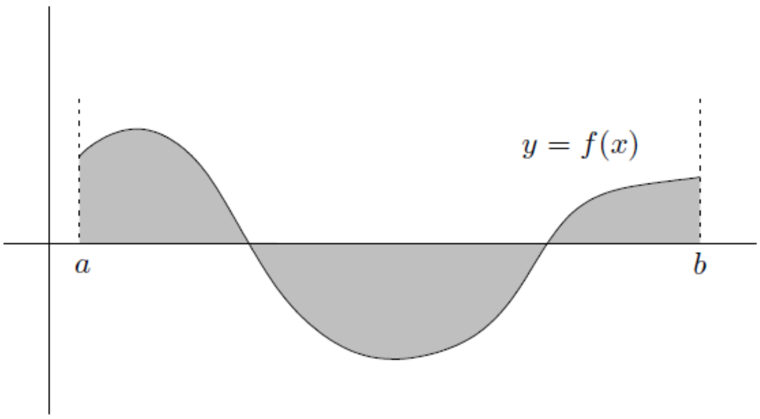


Let's say that the area of the shaded region above, in square units, is

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

This is a **definite** integral. You would read it out loud as "the integral from a to b of $f(x)$ with respect to x." The expression $f(x)$ is called the integrand. The a and b tell you where the two vertical lines go, and are called the limits of integration (not to be confused with regular old limits!) or the *endpoints of integration*. Finally, the dx tells you that x is the variable on the horizontal axis. Actually, x is a **dummy** variable.

What if the function dips below the x-axis? The situation could look like this:



In general, the integral gives the total amount of signed area. More precisely,

$\int_a^b f(x)\,dx$ is the signed area (in square units) of the region between the curve $y = f(x)$, the line $x = a$ and $x = b$, and the x-axis.

Note that the integral is a number, but the area is in square units.

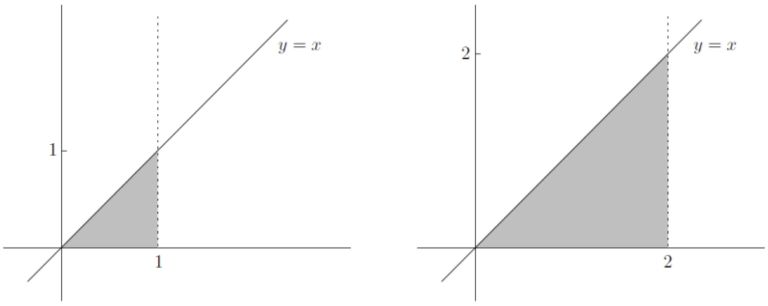
Using our new notation, we can say that

$$\boxed{\text{displacement} = \int_a^b v(t)\,dt} \quad \text{and} \quad \boxed{\text{distance} = \int_a^b |v(t)|\,dt}$$

● Some easy examples

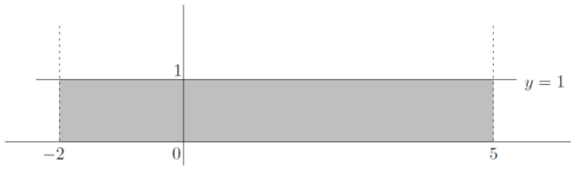
Now, let's look at a few simple examples of definite integrals. First, consider

$$\int_0^1 x dx \quad \text{and} \quad \int_0^2 x dx$$



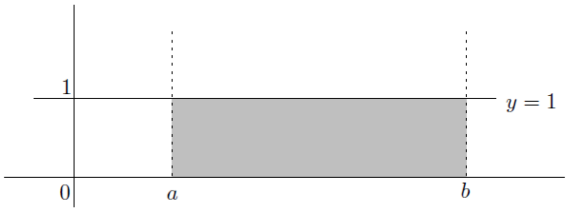
Now, let's take a look at another definite integral:

$$\int_{-2}^{5} 1 dx$$



In fact, the more general integral

$$\int_a^b 1 dx$$

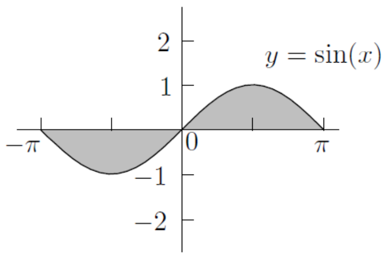


in general. This could also be written as simply

$$\int_a^b dx = b - a$$

Finally, what is

$$\int_{-\pi}^{\pi} \sin(x)\, dx\, ?$$



Before we move on, I'd like to point out a generalization of the previous example. That is

$$\text{if } f \text{ is an odd function, then} \quad \int_{-a}^{a} f(x) dx = 0 \quad \text{for any } a$$

This is true by symmetry: every bit of area *above* the x-axis has a corresponding bit of area below the x-axis, just as in the above picture.

16.2 Definition of the Definite Integral

Unfortunately, the definition of the definite integral is a lot nastier than the above definition of the derivative. The good news is that we've already done the grunt work in the previous chapter, and we can just state the definition:

$$\int_a^b f(x)\,dx = \lim_{\text{mesh}\to 0} \sum_{j=1}^{n} f(c_j)(x_j - x_{j-1})$$

$$\text{where } a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b \text{ and } c_j \text{ is in } [x_{j-1}, x_j] \text{ for each } j = 1, \cdots, n$$

Even though that definition is wordy, it still doesn't tell the full story! (Page 330)

The sum

$$\sum_{j=1}^{n} f(c_j)(x_j - x_{j-1})$$

which appears in the definition is called a *Riemann sum*.

● An example of using the definition

16.3 Properties of Definite Integrals

if you reverse the limits of integration, you need to put in a minus sign out front. In general, for an integrable function f and numbers a and b, we have

$$\boxed{\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx}$$

Now, what if the limits of integration are equal? In fact, it's generally true that

$$\boxed{\int_a^a f(x)\,dx = 0}$$

In general, for any integrable function f and numbers a, b, and c, we have

$$\boxed{\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx}$$

You can split an integral into two pieces, even if the break point c is outside the original interval $[a,b]$, as long as in both pieces the integrand f is still integrable.

There are two more simple properties of integrals which are even more useful. The first is that **constants move through integral signs**. That is, for any integrable f and numbers a, b, and C,

$$\boxed{\int_a^b Cf(x)\,dx = C\int_a^b f(x)\,dx}$$

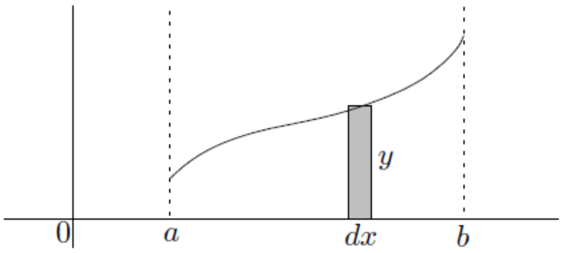
The second property is that **integrals respect sums and differences**. That is, if f and g are both integrable functions, and a and b are two numbers, then

$$\boxed{\int_a^b (f(x)+g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx}$$

16.4 Finding Areas

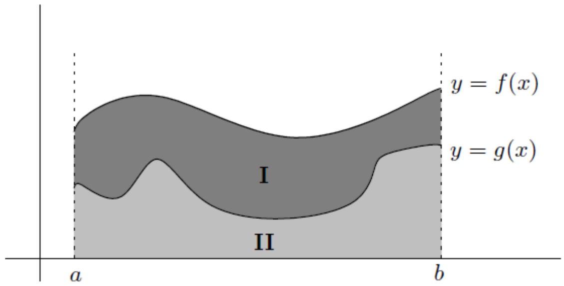
If $y = f(x)$, then we can write

$$\int_a^b y dx$$



This idea is useful in helping to understand how to use the integral to find areas.

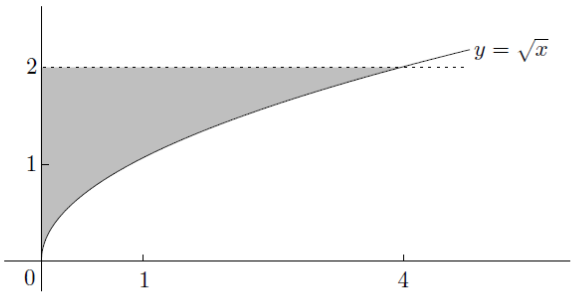
- Finding the unsigned area
- Finding the area between two curves



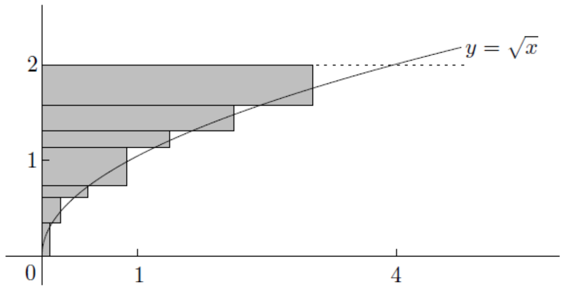
$$\boxed{\text{area between } f \text{ and } g \text{ (in square units)} = \int_a^b |f(x) - g(x)|\,dx}$$

- Finding the area between a curve and the y-axis

Let's try to find the area of the region enclosed by the curve $y = \sqrt{x}$, the y-axis, and the line $y = 2$. Here's a picture of the region:



When we do this, we're effectively chopping up the region we want into horizontal strips, not the vertical ones we've used before. Here's an example of how this might look:



If $y = f(x)$, then $x = f^{-1}(y)$, provided that the inverse function exists. So, we can summarize the situation as follows:

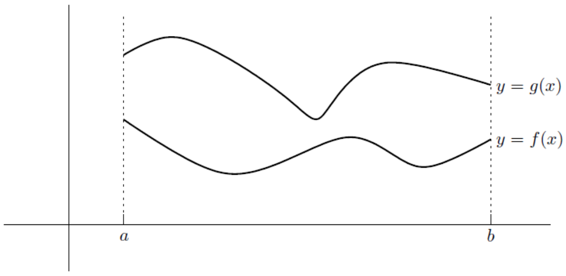
> $\int_A^B f^{-1}(y)\,dy$ is the signed area (in square units) of the region between the curve $y = f(x)$, the lines $y = A$ and $y = B$, and the y-axis, if f is invertible.

If you prefer, you can write the above integral as

$$\int_A^B x\,dy$$

16.5 Estimating Integrals

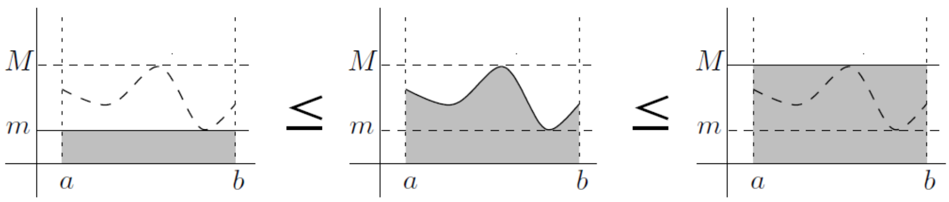
Here's a very simple but important principle: **when one function is always larger than another, its integral is also larger**. Take a look at the following picture:



On the interval $[a,b]$, the function g always lies above f. In any case, the area under $y = f(x)$ (down to the x-axis) is clearly less than the area under $y = g(x)$ (down to the x-axis). In symbols:

> if $f(x) \leq g(x)$ for all x in $[a,b]$, then $\int_a^b f(x)dx \leq \int_a^b g(x)dx$

● A simple type of estimation



> if $m \leq f(x) \leq M$ for all x in $[a,b]$, then $m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)$

16.6 Averages and the Mean Value Theorem for Integrals

If the time interval goes from a to b, and the velocity at time t is $v(t)$, then we've already seen that

$$\text{displacement} = \int_a^b v(t)dt$$

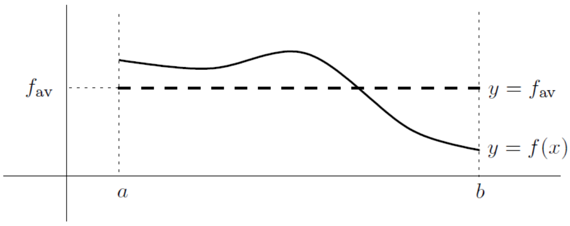
Since the total time is $b-a$, we have

$$\text{average velocity} = \frac{\text{displacement}}{\text{total time}} = \frac{1}{b-a}\int_a^b v(t)dt$$

More generally, we can define the average value of an integrable function f on the interval $[a,b]$ as follows:

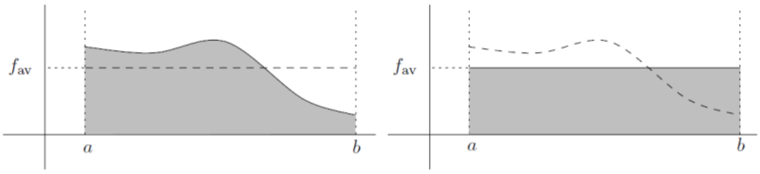
$$\boxed{\text{average value of } f \text{ on } [a,b] = \frac{1}{b-a}\int_a^b f(x)dx}$$

For example,



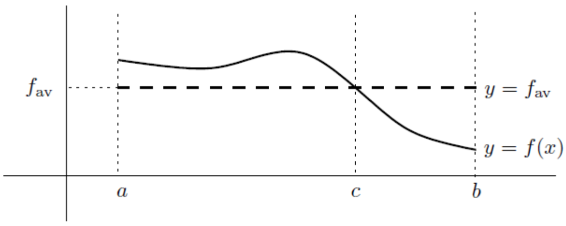
$$f_{\text{av}} = \frac{1}{b-a}\int_a^b f(x)dx$$

This actually says that the following two areas are equal:



● The Mean Value Theorem for integrals

In the above graphs, observe that the horizontal line $y = f_{\text{av}}$ intersects the graph of $y = f(x)$. Let's label the corresponding point on the x-axis as c, like this:



So we have $f(c) = f_{\text{av}}$. It turns out that if f is continuous, then there is always such a number c:

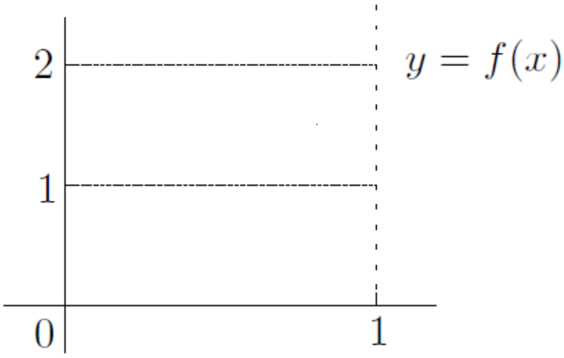
Mean Value Theorem for integrals: if f is continuous on $[a,b]$, then there exists c in (a,b) such that $f(c) = \frac{1}{b-a}\int_a^b f(x)dx$.

So, why is the above theorem also called the Mean Value Theorem? The difference between the two versions of the theorem is that in the regular version, the conclusion was interpreted in terms of slopes on the graph of displacement versus time; whereas now we have interpreted it in terms of areas on the graph of velocity versus time.

16.7 A Nonintegrable Function

Now, let's look at an example of a function where there **are** too many discontinuities

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 2 & \text{if } x \text{ is irrational} \end{cases}$$



CHAPTER 17 The Fundamental Theorems of Calculus

17.1 Functions Based on Integrals of Other Functions

Let's try to find

$$\int_0^{\text{any number}} x^2 dx$$

There are two ways we can make this whole thing more general. First, the left-hand endpoint doesn't have to be 0

$$F(x) = \int_0^x t^2 dt$$

$$G(x) = \int_2^x t^2 dt$$

$$F(x) = \frac{8}{3} + G(x)$$

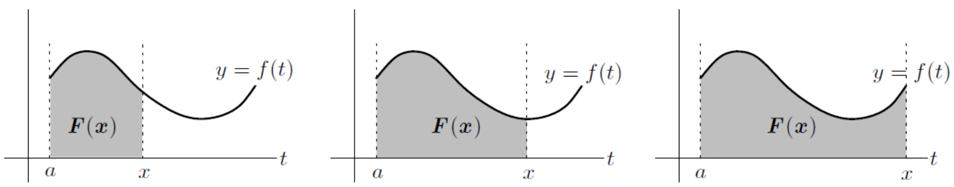
$$H(x) = \int_a^x t^2 dt$$

$$F(x) = H(x) + C$$

The moral of the story is that changing the left-hand endpoint from one constant to another doesn't make too much difference.

Our second generalization is that the integrand doesn't have to be t^2. It can be any continuous function of t. Let's suppose the integrand is $f(t)$. If a is some constant number, then let's define

$$F(x) = \int_a^x f(t)dt$$



17.2 The First Fundamental Theorem

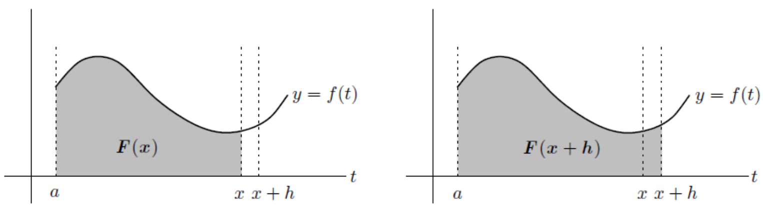
Here's the goal: find

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

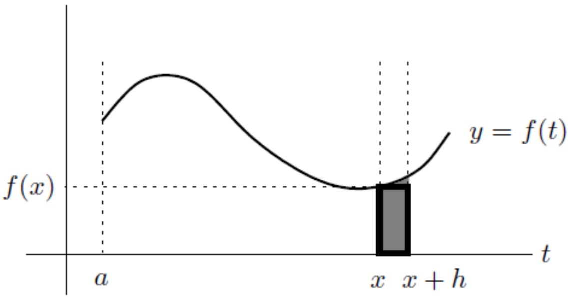
without using Riemann sums. Let's do three things which are not really obvious at all (Page 358)

$$F(x) = \int_a^x f(t)dt$$

$$F'(x) = \frac{d}{dx}\int_a^x f(t)dt$$



$$F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h}$$



$$\lim_{h\to 0}\frac{F(x+h)-F(x)}{h}=f(x)$$

First Fundamental Theorem of Calculus: for f continuous on $[a,b]$, define a function F by

$$F(x)=\int_a^x f(t)\,dt \quad \text{for } x \text{ in } [a,b]$$

Then F is differentiable on (a,b) and $F'(x)=f(x)$

● Introduction to antiderivatives

Suppose that $f(t)=t^2$ and $a=0$, so that

$$F(x)=\int_0^x t^2dt$$

The First Fundamental Theorem tells us that $F'(x)=f(x)$. Since $f(t)=t^2$, we have $f(t)=x^2$, this means that $F'(x)=x^2$. In other words, F is a function whose derivative is x^2. We say that F is an ***antiderivative*** of x^2 (with respect to x).

17.3 The Second Fundamental Theorem

The example with $f(t)=t^2$ in the previous section points the way to finding $\int_a^b f(t)dt$ in general. First, we know that the function F defined by

$$F(x)=\int_a^x f(t)dt$$

is an antiderivative of f (with respect to x). We really want to find $F(b)$ (Page 362)

Second Fundamental Theorem of Calculus: if f is continuous on $[a,b]$, and F is any antiderivative of f (with respect to x), then

$$\int_a^b f(x)\,dx=F(b)-F(a)$$

Then F is differentiable on (a,b) and $F'(x)=f(x)$

In practice, the right-hand side is normally written as $F(x)\Big|_a^b$. That is, we set

$$F(x)\Big|_a^b=F(b)-F(a)$$

17.4 Indefinite Integrals

We might as well have a shorthand way of expressing antiderivatives without having to write the long word “antiderivative.” Inspired by the First Fundamental Theorem, we'll write

$$\int f(x)dx$$

to mean "the family of all antiderivatives of f." Bear in mind that any integrable function has infinitely many antiderivatives, but they all differ by a constant. This is what I mean when I say "family." For example,

$$\int x^2 dx = \frac{x^3}{3} + C$$

for some constant C.

If you know a derivative, you get an antiderivative for free. In particular:

$$\boxed{\text{if} \quad \frac{d}{dx}F(x) = f(x), \quad \text{then} \quad \int f(x)dx = F(x) + C}$$

The above example fits this pattern:

$$\frac{d}{dx}\left(\frac{x^3}{3}\right) = x^2, \quad \text{so} \quad \int x^2 dx = \frac{x^3}{3} + C$$

$$\frac{d}{dx}(\sin(x)) = \cos(x), \quad \text{so} \quad \int \cos(x)\, dx = \sin(x) + C$$

$$\frac{d}{dx}(\tan^{-1}(x)) = \frac{1}{1+x^2}, \quad \text{so} \quad \int \frac{1}{1+x^2} dx = \tan^{-1}(x) + C$$

All the above integrals are examples of ***indefinite integrals***. You can tell an indefinite integral from a definite integral by noticing whether or not there are limits of integration. Indefinite integrals don't have limits of integration, while definite integrals do:

● A definite integral, like $\int_a^b f(x)dx$, is a **number**. It represents the signed area of the region bounded by the curve $y = f(x)$, the x-axis, and the lines $x = a$ and $x = b$

● An indefinite integral, like $\int f(x)dx$, is a **family of functions**. This family consists of all functions which are antiderivatives of f (with respect to x). The functions all differ by a constant.

Here are two simple facts about indefinite integrals that follow directly from the similar properties for derivatives: if f and g are integrable, and c is a constant, then

$$\boxed{\int \big(f(x) + g(x)\big)dx = \int f(x)dx + \int g(x)dx}$$

and

$$\boxed{\int cf(x)dx = c\int f(x)dx}$$

That is, the integral of the sum is the sum of the integrals, and constant multiples can be pulled through the integral sign.

I want to make one more comment about the two Fundamental Theorems. The First Fundamental Theorem says that

$$\frac{d}{dx}\int_a^x f(t)dt = f(x)$$

In some sense, the derivative of the integral is the original function. You just have to be careful about what you mean by the "integral," bearing in mind that the variable has to be the right-hand limit of integration, not the dummy variable. Now, the Second Fundamental Theorem says that

$$\int_a^b f(x)dx = F(x)\Big|_a^b$$

where F is an antiderivative of f. This means that $f(x) = \frac{d}{dx}F(x)$. We can therefore rewrite the above equation as

$$\int_a^b \frac{d}{dx}F(x)dx = F(x)\Big|_a^b$$

which can be interpreted as saying that the integral of the derivative is the original function. Again, it's not really the original function: it's the difference between the evaluations of the original function at the endpoints a and b. Even with all this vagueness, it should still be clear that differentiation and integration are essentially opposite operations.

17.5 How to Solve Problems: The First Fundamental Theorem

Think about how you'd find the following derivative:

$$\frac{d}{dx}\int_3^x \sin(t^2)\,dt$$

Why go to all that work when the derivative and integral effectively cancel each other out? After all, if you wanted to find $\left(\sqrt{54756}\right)^2$, you wouldn't waste time looking for $\sqrt{54756}$ when you just have to square it again. You'd just write down the answer 54756 and be done with it. Similarly, we can use the First Fundamental Theorem from above to say that

$$\frac{d}{dx}\int_3^x \sin(t^2)\,dt = \sin(x^2)$$

All you have to do is take the integrand $\sin(t^2)$ and change t to x. The number 3 doesn't even come into it. By the way, it would be a mistake to put a "$+C$" at the end: you are finding a **derivative**, after all, not an antiderivative!

● Variation 1: variable left-hand limit of integration

Consider

$$\frac{d}{dx}\int_x^7 t^3\cos(t\ln(t))\,dt$$

The problem is that the variable x is now the left-hand limit of integration, not the right-hand one we've been used to. No problem-just switch the x and 7 around, introducing a minus sign to compensate for this. You get

$$\frac{d}{dx}\int_x^7 t^3\cos(t\ln(t))\,dt = \frac{d}{dx}\left(-\int_7^x t^3\cos(t\ln(t))\,dt\right)$$

● Variation 2: one tricky limit of integration

Here's another example:

$$\frac{d}{dx}\int_0^{x^2} \tan^{-1}(t^7+3t)\,dt$$

Because the right-hand limit of integration is x^2, not x, we can't just use the First Fundamental Theorem directly. We're going to need the chain rule as well

$$y = \int_0^{x^2} \tan^{-1}(t^7+3t)\,dt$$

$$y = \int_0^{u} \tan^{-1}(t^7+3t)\,dt$$

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

$$\frac{dy}{du} = \frac{d}{du}\int_0^u \tan^{-1}(t^7 + 3t)\,dt = \tan^{-1}(u^7 + 3u)$$

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = (\tan^{-1}(u^7 + 3u))(2x)$$

$$\frac{dy}{dx} = 2x\tan^{-1}(x^{14} + 3x^2)$$

In summary,

$$\frac{d}{dx}\int_0^{x^2} \tan^{-1}(t^7 + 3t)\,dt = 2x\tan^{-1}(x^{14} + 3x^2)$$

Let's look at one more example of this sort of problem: what is

$$\frac{d}{dq}\int_4^{\sin(q)} \tan(\cos(a))\,da$$

$$\frac{dy}{dq} = \frac{dy}{du}\frac{du}{dq} = \tan(\cos(u))\cos(q) = \tan(\cos(\sin(q)))\cos(q)$$

You might also encounter both of the above variations in the same problem. For example, to find

$$\frac{d}{dq}\int_{\sin(q)}^{4} \tan(\cos(a))\,da$$

● Variation 3: two tricky limits of integration

Here's an even harder example:

$$\frac{d}{dx}\int_{x^5}^{x^6} \ln(t^2 - \sin(t) + 7)\,dt$$

Now there are functions of x in both the left-hand and right-hand limits of integration. The way to handle this is to split the integral into two pieces at some number. It actually doesn't matter where you split it, as long as it is at a constant (where the function is defined). So, pick your favorite number-say 0-and split the integral there:

$$\frac{d}{dx}\int_{x^5}^{x^6} \ln(t^2 - \sin(t) + 7)\,dt$$

$$= \frac{d}{dx}\left(\int_{x^5}^{0} \ln(t^2 - \sin(t) + 7)\,dt + \int_0^{x^6} \ln(t^2 - \sin(t) + 7)\,dt\right)$$

$$= -5x^4\ln(x^{10} - \sin(x^5) + 7) + 6x^5\ln(x^{12} - \sin(x^6) + 7)$$

● Variation 4: limit is a derivative in disguise

Here's an example which looks a little different:

$$\lim_{h\to 0}\frac{1}{h}\int_x^{x+h} \log_3(\cos^6(t) + 2)\,dt$$

$$F(x) = \int_a^x \log_3(\cos^6(t) + 2)\,dt$$

$$F(x+h) - F(x) = \int_x^{x+h} \log_3(\cos^6(t) + 2)\,dt$$

$$\lim_{h\to 0}\frac{1}{h}\int_x^{x+h} \log_3(\cos^6(t) + 2)\,dt = \lim_{h\to 0}\frac{F(x+h) - F(x)}{h} = F'(x)$$

So actually, we have

$$\lim_{h\to 0}\frac{1}{h}\int_x^{x+h}\log_3(\cos^6(t)+2)\,dt=\frac{d}{dx}\int_a^x\log_3(\cos^6(t)+2)\,dt$$

for any a you like. See, I told you that the limit was a derivative in disguise! To finish the problem, just apply the First Fundamental Theorem in its basic form to see that the above limit is just $\log_3(\cos^6(x)+2)$.

17.6 How to Solve Problems: The Second Fundamental Theorem

To find a definite integral using the Second Fundamental Theorem.

● Finding indefinite integrals

Start off by noting that

$$\frac{d}{dx}(x^{a+1})=(a+1)x^a$$

this means that

$$\int (a+1)x^a dx = x^{a+1}+C$$

If $a\neq -1$, then $a+1\neq 0$; so we can divide through by $(a+1)$ and write

$$\boxed{\int x^a dx=\frac{x^{a+1}}{a+1}+C}$$

(Once again, we replaced $C/(a+1)$ by simply C; this is OK since C is just an arbitrary constant.) Now, what happens when $a=-1$? The above method doesn't work on $\int x^{-1}dx$, which is just

$$\int \frac{1}{x}dx$$

we can prove the formula

$$\boxed{\int \frac{1}{x}dx=\ln|x|+C}$$

In the meantime, we can now summarize most of the basic derivatives and corresponding antiderivatives that we've seen so far in one big table

Derivatives and integrals to learn:

$\frac{d}{dx}x^a=ax^{a-1}$	$\int x^a dx=\frac{x^{a+1}}{a+1}+C \quad (\text{if } a\neq -1)$
$\frac{d}{dx}\ln(x)=\frac{1}{x}$	$\int \frac{1}{x}\,dx=\ln\lvert x\rvert+C$
$\frac{d}{dx}e^x=e^x$	$\int e^x dx=e^x+C$
$\frac{d}{dx}b^x=b^x\ln(b)$	$\int b^x dx=\frac{b^x}{\ln(b)}+C$
$\frac{d}{dx}\sin(x)=\cos(x)$	$\int \cos(x)\,dx=\sin(x)+C$
$\frac{d}{dx}\cos(x)=-\sin(x)$	$\int \sin(x)\,dx=-\cos(x)+C$
$\frac{d}{dx}\tan(x)=\sec^2(x)$	$\int \sec^2(x)\,dx=\tan(x)+C$

$$\frac{d}{dx}\sec(x) = \sec(x)\tan(x) \qquad \int \sec(x)\tan(x)\,dx = \sec(x) + C$$

$$\frac{d}{dx}\cot(x) = -\csc^2(x) \qquad \int \csc^2(x)\,dx = -\cot(x) + C$$

$$\frac{d}{dx}\csc(x) = -\csc(x)\cot(x) \qquad \int \csc(x)\cot(x)\,dx = -\csc(x) + C$$

$$\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{1-x^2}} \qquad \int \frac{1}{\sqrt{1-x^2}}dx = \sin^{-1}(x) + C$$

$$\frac{d}{dx}\tan^{-1}(x) = \frac{1}{1+x^2} \qquad \int \frac{1}{1+x^2}dx = \tan^{-1}(x) + C$$

$$\frac{d}{dx}\sec^{-1}(x) = \frac{1}{|x|\sqrt{x^2-1}} \qquad \int \frac{1}{|x|\sqrt{x^2-1}}dx = \sec^{-1}(x) + C$$

$$\frac{d}{dx}\sinh(x) = \cosh(x) \qquad \int \cosh(x)\,dx = \sinh(x) + C$$

$$\frac{d}{dx}\cosh(x) = \sinh(x) \qquad \int \sinh(x)\,dx = \cosh(x) + C$$

As we've seen, if you replace x by the constant multiple ax in any of the above differentiation formulas, you just have to multiply the corresponding formula by a. For example,

$$\frac{d}{dx}\tan(7x) = 7\sec^2(7x)$$

What if you integrate instead? Now the rule of thumb is that if you replace x by ax, then you have to **divide** by a. For example,

$$\int \sec^2(7x)\,dx = \frac{1}{7}\tan(7x) + C$$

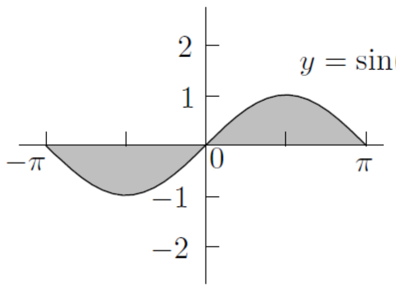
● Finding definite integrals

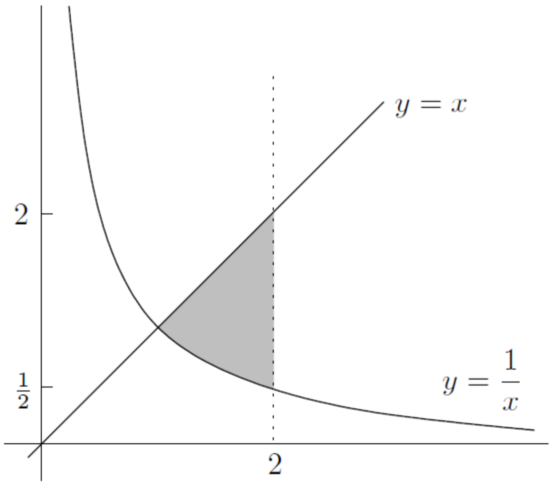
The Second Fundamental Theorem tells us that to find

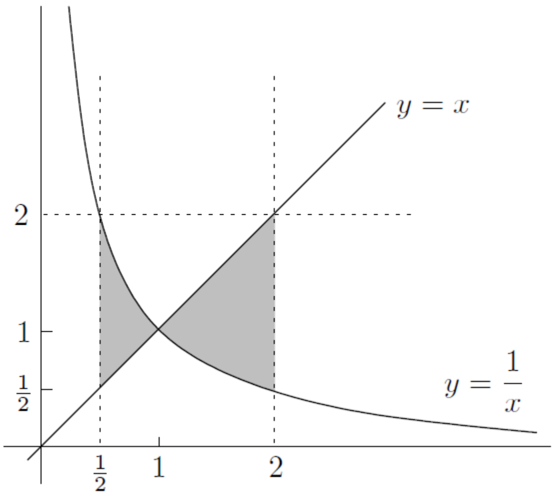
$$\int_a^b f(x)dx$$

just find an antiderivative, plug in $x = b$ and $x = a$, and take the difference (Page 374)

● Unsigned areas and absolute values





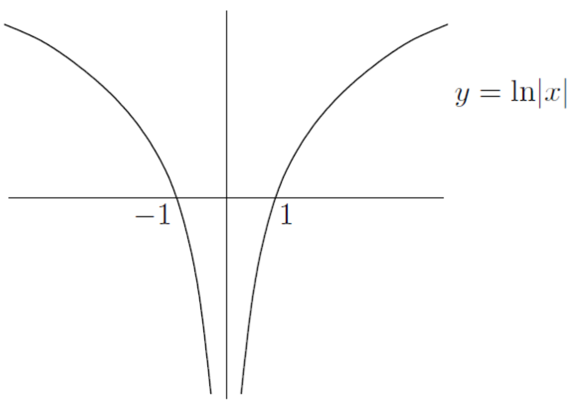


17.7 A Technical Point

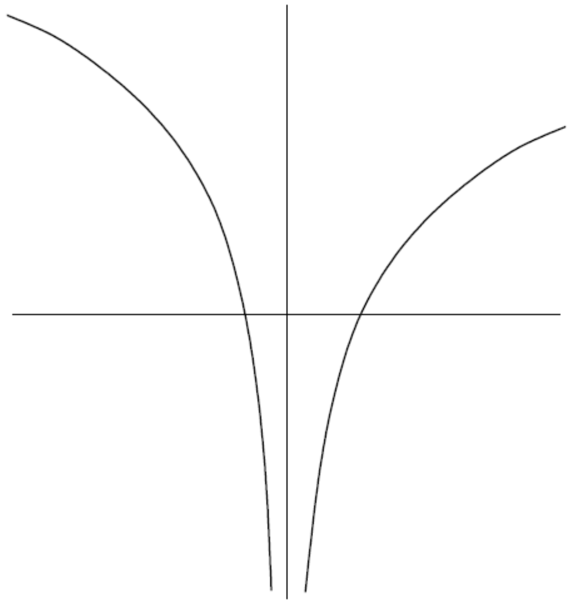
We saw that

$$\int \frac{1}{x} dx = \ln|x| + C$$

Although everyone writes the formula like this, technically it's not correct! To see why, let's start off with the graph of $y = \ln|x|$:



This has two pieces, either of which can be shifted up or down without affecting the derivative. For example, if we shift the left piece up by 1 and the right piece down by $1/2$, the graph looks something like this:



This function isn't of the form $\ln|x| + C$, but its derivative is still $1/x$. So we really need to allow **two** constants, possibly different-one for each of the two pieces of the curve:

$$\int \frac{1}{x} dx = \begin{cases} \ln|x| + C_1 & \text{if } x < 0 \\ \ln|x| + C_2 & \text{if } x > 0 \end{cases}$$

But there's a problem: the vertical asymptote at $x = 0$. So the only time that definite integrals of the form

$$\int_a^b \frac{1}{x} dx$$

make sense is when a and b are both positive or both negative. In either case, only one of the pieces of $\ln|x|$ is involved, and there's no need to mess around with two different constants!

17.8 Proof of the First Fundamental Theorem

(Page 381)

18.1 Substitution

Actually, this is a special case of a nice fact: if f is a differentiable function, then

$$\boxed{\int \frac{f'(x)}{f(x)}dx = \ln|f(x)| + C}$$

So if the top is the derivative of the bottom, then the integral is just the log of the bottom (with absolute values and the $+C$.

● Substitution and definite integrals

You can also use the substitution method on definite integrals. There are two legitimate ways to do this. For example, to find

$$\int_0^{\sqrt[3]{\pi/2}} x^2\cos(x^3)\,dx$$

you could find the indefinite integral $\int x^2\cos(x^3)\,dx$ first, then plug in the limits of integration

$$\int x^2\cos(x^3)\,dx = \int \cos(t)\frac{dt}{3} = \frac{1}{3}\int \cos(t)\,dt = \frac{1}{3}\sin(t) + C = \frac{1}{3}\sin(x^3) + C$$

It's really important to go back to x-land at the last step. Now we can use the Second Fundamental Theorem

$$\int_0^{\sqrt[3]{\pi/2}} x^2\cos(x^3)\,dx = \frac{1}{3}\sin(x^3)\Big|_0^{\sqrt[3]{\pi/2}} = \left(\frac{1}{3}\sin\left(\left(\sqrt[3]{\pi/2}\right)^3\right)\right) - \left(\frac{1}{3}\sin(0^3)\right)$$

which works out to be $\frac{1}{3}$. So one way to use the substitution method on a definite integral is to focus on the indefinite integral first, then after you've found it, plug in the limits of integration. We'll finish this soon, but first note that it would be a major error to write

$$\frac{1}{3}\int_0^{\sqrt[3]{\pi/2}} \cos(t)\,dt$$

on the right-hand side instead. Since we're integrating with respect to t, not x, the limits of integration must refer to relevant values of t

$$\int_{x=0}^{x=\sqrt[3]{\pi/2}} x^2\cos(x^3)\,dx = \frac{1}{3}\int_{t=0}^{t=\pi/2} \cos(t)\,dt$$

So, all in all, we've substituted **three** things:

1. the dx bit-that became something to do with dt, burning up some of the other x stuff in order to make the change;
2. all the remaining terms in the integrand involving x, so that they became terms in t;
3. the limits of integration.

The best way to set it out is to make a working column at the left of your page, like this (Page 387).

● How to decide what to substitute

How do you choose the substitution? Good question. The basic idea is to look for some component of the integrand whose derivative is also present as a factor of the integrand. In the integral

$$\int \frac{1}{\sin^{-1}(x)\sqrt{1-x^2}}dx$$

the substitution $t=\sin^{-1}(x)$ works because its derivative $1/\sqrt{1-x^2}$ is right there, waiting for us to use it.
Sometimes the substitution is not obvious at all (Page 390)
Let's look at one more example:

$$\int x\sqrt[5]{3x+2}dx$$

There is a nice technique for dealing with integrals involving terms such as $\sqrt[n]{ax+b}$. You simply set $t=\sqrt[n]{ax+b}$, but take nth powers before you differentiate to find dt. So:

to deal with $\sqrt[n]{ax+b}$, set $t=\sqrt[n]{ax+b}$ and differentiate both sides of $t^n=ax+b$

By the way, did you notice anything different about this substitution from all the others we've done so far? It's a subtle point, but in all the other examples, we had an equation like $dt=(x-\text{stuff})dx$, whereas here, we have $dx=(t-\text{stuff})dt$. This worked out quite nicely, since we just replaced dx directly. In all the other examples, we had to find a constant multiple of the x-stuff already present in order to have much of a chance.
Addition:

(1) $\iint_D \cos[\tan^{-1}(x^2+y^2)]\,dx\,dy,\ D=\{(x,y)|x^2+y^2\le\sqrt{3}\}$

$$\begin{matrix} x=r\cos\theta \\ y=r\sin\theta\end{matrix},\quad 0\le r\le 3^{\frac{1}{4}},\quad 0\le\theta\le 2\pi$$

$$\iint_D \cos[\tan^{-1}(x^2+y^2)]\,dx\,dy=\iint_D \cos[\tan^{-1}(r^2)]\,r\,dr\,d\theta=\int_0^{2\pi}\int_0^{3^{\frac{1}{4}}}\cos[\tan^{-1}(r^2)]\,r\,dr\,d\theta$$

$$=2\pi\int_0^{3^{\frac{1}{4}}}\cos[\tan^{-1}(r^2)]\,r\,dr$$

Substitution: $\tan^{-1}(r^2)=t$, so $\tan(t)=r^2$, $\frac{dr^2}{dt}=2r\frac{dr}{dt}=\sec^2(t)\to dr=\frac{\sec^2(t)}{2r}dt$.

Then,

$$\iint_D \cos[\tan^{-1}(x^2+y^2)]\,dx\,dy=$$

$$2\pi\int_0^{3^{\frac{1}{4}}}\cos[\tan^{-1}(r^2)]\,r\,dr=2\pi\int_0^{\frac{\pi}{3}}\cos(t)\frac{\sec^2(t)}{2}dt=\pi\int_0^{\frac{\pi}{3}}\sec(t)\,dt=\pi\ln\left(2+\sqrt{3}\right).$$

Or another way:

$$\begin{matrix} x=r\cos\theta \\ y=r\sin\theta\end{matrix},\quad 0\le r\le 3^{\frac{1}{4}},\quad 0\le\theta\le 2\pi$$

$$\iint_D \cos[\tan^{-1}(x^2+y^2)]\,dx\,dy=2\pi\int_0^{3^{\frac{1}{4}}}\cos[\tan^{-1}(r^2)]\,r\,dr$$

Since $\tan^2(\theta)+1=\sec^2(\theta)=\frac{1}{\cos^2(\theta)}$, we can get $\cos^2(\theta)=\frac{1}{\tan^2(\theta)+1}$. So

$$\cos(\theta) = \frac{1}{\sqrt{\tan^2(\theta)+1}}$$

Then,

$$2\pi \int_0^{3^{\frac{1}{4}}} \cos[\tan^{-1}(r^2)]\, r\, dr = 2\pi \int_0^{3^{\frac{1}{4}}} \frac{1}{\sqrt{\tan^2(\tan^{-1}(r^2))+1}} r\, dr = 2\pi \int_0^{3^{\frac{1}{4}}} \frac{1}{\sqrt{r^4+1}} r\, dr$$

$$= 2\pi \times \frac{1}{2}\int_0^{3^{\frac{1}{4}}} \frac{1}{\sqrt{r^4+1}} dr^2 = \pi \int_0^{3^{\frac{1}{2}}} \frac{1}{\sqrt{t^2+1}} dt$$

Using $t = \tan(\alpha)$, so

$$\pi \int_0^{3^{\frac{1}{2}}} \frac{1}{\sqrt{t^2+1}} dt = \pi \int_0^{\frac{\pi}{3}} \frac{\sec^2(\alpha)}{\sec(\alpha)} d\alpha = \pi \int_0^{\frac{\pi}{3}} \sec(\alpha)\, d\alpha = \pi \ln\left(2+\sqrt{3}\right).$$

In general, there are no hard and fast rules about what to substitute. You just have to go along with your instinct, which will be accurate only if you have done plenty of practice problems. You can always try any substitution you like. If the new integral is worse than the original one, or you can't see how to migrate everything to t-land, then don't panic: just go back to the original integral and try something else.

There are two things I want to deal with. One is a justification of the substitution method; I'll do this in the next section. The other is to summarize the method of substitution:

1. for **indefinite** integrals, change everything to do with x and dx to stuff involving t and dt, do the new integral, then change back to x stuff;

2. for **definite** integrals, change everything to do with x and dx to stuff involving t and dt, and change the limits of integration to the corresponding t values as well, then do the new integral (no need to go back to x-land here). Alternatively, treat the integral as an indefinite integral and when you get the final answer, then substitute in the limits of integration.

● Theoretical justification of the substitution method

(Page 392)

A good way to think of $dt/dx = 2x$ or $dt = 2xdx$ is that a change in x produces a change in t which is $2x$ times as large.

the t-land part of the calculation looks like this:

$$\int f(t)dt = F(t) + C$$

We know that $F'(t) = f(t)$. Since $t = g(x)$, we have $F'\big(g(x)\big) = f\big(g(x)\big)$. The above equation becomes

$$\int f\big(g(x)\big)g'(x)dx = F\big(g(x)\big) + C$$

By the way, this nice equation allows us to prove the alternative method of substitution. In the alternative method, instead of setting $t = g(x)$, we set $x = g(t)$ for some other function g, and replaced dx by $g'(t)dt$. In that case, our original integral $\int f(x)dx$ now supposedly becomes

$$\int f\big(g(t)\big)g'(t)dt$$

18.2 Integration by Parts

We saw how to reverse the chain rule by using the method of substitution. There is also a way to reverse the product rule-it's called integration by parts. Let's recall the product rule

$$\frac{d}{dx}(uv) = v\frac{du}{dx} + u\frac{dv}{dx}$$

Let's rearrange this equation and then integrate both sides with respect to x. We get

$$\int u\frac{dv}{dx}dx = \int \frac{d}{dx}(uv)dx - \int v\frac{du}{dx}dx$$

$$\int u\frac{dv}{dx}dx = uv - \int v\frac{du}{dx}dx$$

It's perfectly usable in this form, but there's an abbreviated form which is even more convenient.

If we replace $\frac{dv}{dx}dx$ by dv, and replace $\frac{du}{dx}dx$ by du, we get the formula

$$\boxed{\int udv = uv - \int vdu}$$

Let's see how it works in practice. Suppose we want to find

$$\int xe^x dx$$

Substitution seems useless (try it and see), so let's try integration by parts. Set $u = x$ and $dv = e^x dx$. I recommend writing the following:

$$u = x \qquad\qquad v =$$

$$du = \qquad\qquad dv = e^x dx$$

and then filling in the blanks by differentiating u and integrating dv:

$$u = x \qquad\qquad v = e^x$$

$$du = dx \qquad\qquad dv = e^x dx$$

The easiest way to use the formula is to write a small version of it with generous spacing, then do the substitutions underneath, like this:

$$\int u \quad dv \quad = u \ v \ - \int \ v \quad du$$

$$\int x \quad \overbrace{e^x dx} = x \ e^x \ - \int \ e^x \quad dx$$

Now we still have one integral left, but it's just $\int e^x\,dx$, which is $e^x + C$. Plugging this in, we see that $\int xe^x\,dx = xe^x - e^x + C$ (Technically it should be $-C$, not $+C$, but minus a constant is just another constant and there's no need to distinguish.)

Now, how on earth did we know to choose $u = x$ and $dv = e^x dx$? Why couldn't we have chosen $u = e^x$ and $dv = xdx$? Well, we could have. But it's not very useful. The moral is that if e^x is present, you should normally let $dv = e^x dx$ so that v is simply equal to e^x.

● Some variations

A few complications can arise. Sometimes you need to integrate by parts twice or more. For example, how would you find (Page 394)

$$\int x^2 \sin(x)\,dx$$

Sometimes you can integrate by parts twice but things don't seem to get simpler. In this case, if you're lucky, then you might just get a multiple of the original integral back at the end. Then unless you are actually unlucky, you can throw it over to the other side and solve, which is a neat trick. (If you are unlucky, then the integrals cancel out, which doesn't help at all!) (Page 395)

There's one other type of integral that needs integration by parts but is in disguise. In particular, the integrand doesn't appear to be a product. Some integrals that fall into this category are

$$\int \ln(x)\,dx, \quad \int (\ln(x))^2 dx, \quad \int \sin^{-1}(x)\,dx, \quad \text{and} \quad \int \tan^{-1}(x)\,dx$$

That is, the integrand is any inverse trig function (by itself) or a power of $\ln(x)$. In this case, you should let u be the integrand itself, and let $dv = dx$ (Page 397).

When solving a definite integral by integrating by parts, find the indefinite integral first, then substitute the limits of integration at the end.

18.3 Partial Fractions

Let's focus our attention on how to integrate a rational function. So we want to find an integral like

$$\int \frac{p(x)}{q(x)} dx$$

where p and q are polynomials. This covers a whole slew of integrals, for example,

$$\int \frac{x^2+9}{x^4-1} dx, \quad \int \frac{x}{x^3+1} dx, \quad \text{or} \quad \int \frac{1}{x^3-2x^2+3x-7} dx$$

These seem a little complicated. Here are some simpler ones:

$$\int \frac{1}{x-3} dx, \quad \int \frac{1}{(x+5)^2} dx, \quad \int \frac{1}{x^2+9} dx, \quad \text{and} \quad \int \frac{3x}{x^2+9} dx$$

The last four integrands are all rational functions, but they are a lot simpler. Try to work out all of these integrals using substitution. The first two of these integrals have denominators which are powers of linear functions, whereas the last two have quadratic denominators which cannot be factored.

So, here's the idea: first we'll see how to take a general rational function and do some algebra to bust it up into a sum of simpler rational functions; then we'll see how to integrate the simpler types of rational functions. The simpler functions I'm talking about are all like the four above: they either look like a constant over a linear power, or they look like a linear function over a quadratic.

● The algebra of partial fractions

Our goal is to break up a rational function into simpler pieces. The first step in this process is to make sure that the numerator of the function has degree less than the denominator. If not, we'll have to start off with a long division. So in the examples

$$\int \frac{x+2}{x^2-1} dx \quad \text{and} \quad \int \frac{5x^2+x-3}{x^2-1} dx$$

The first is fine, but the second example isn't so great, because the degrees of the top and bottom are equal (to 2). We'd have the same trouble if there were a cubic or higher-degree polynomial on the top. So, we have to do a long division. To do this, write

$$\text{denominator}\overline{)\text{numerator}}$$

So we have

$$\frac{5x^2+x-3}{x^2-1} = 5 + \frac{x+2}{x^2-1}$$

If we integrate both sides with respect to x, we can break up the integral into two pieces, and actually do the integral in the first piece, to see that our original integral is equal to

$$\int 5dx + \int \frac{x+2}{x^2-1}dx = 5x + \int \frac{x+2}{x^2-1}dx$$

The new integral has a degree of 1 on the top and 2 on the bottom, which is the way we like it. We're now ready to proceed.

Next, we'll factor the denominator. If the denominator is a quadratic, check the discriminant: if this is negative, you can't factor the quadratic. Otherwise, you can factor it by hand or by using the quadratic formula. If your denominator is more complicated, you may have to guess a root and do a long division.

After factoring the denominator, the next step is to write down something called the "form." This is made by adding together one or more terms for each factor of the denominator, according to the following rules:

1. If you have a linear factor $(x+a)$, then the form has a term like

$$\frac{A}{x+a}$$

2. If you have the square of a linear factor $(x+a)^2$, then the form has terms like

$$\frac{A}{(x+a)^2} + \frac{B}{x+a}$$

3. If you have a quadratic factor (x^2+ax+b), then the form has a term like

$$\frac{Ax+B}{x^2+ax+b}$$

Those are the most common ones. Here are some rarer beasts:

4. If you have the cube of a linear factor $(x+a)^3$, then the form has terms like

$$\frac{A}{(x+a)^3} + \frac{B}{(x+a)^2} + \frac{C}{x+a}$$

5. If you have the fourth power of a linear factor $(x+a)^4$, then the form has terms like

$$\frac{A}{(x+a)^4} + \frac{B}{(x+a)^3} + \frac{C}{(x+a)^2} + \frac{D}{x+a}$$

Notice that **the form only depends on the denominator**. The numerator is irrelevant! Also, when I use constants like A, B, C, and D above, bear in mind that you can't reuse constants in different terms.

So you need to keep advancing along the alphabet. Since the denominator factors as $(x-1)(x+1)$, we have two linear factors, and the form is

$$\frac{A}{x-1} + \frac{B}{x+1}$$

We can't use A twice, so we used B for the second term. Here's another example of finding the form. What would the form of

$$\frac{\text{any old junk}}{(x-1)(x+4)^3(x^2+4x+7)(3x^2-x+1)}$$

be? The answer is

$$\frac{A}{x-1} + \frac{B}{(x+4)^3} + \frac{C}{(x+4)^2} + \frac{D}{x+4} + \frac{Ex+F}{x^2+4x+7} + \frac{Gx+H}{3x^2-x+1}$$

Once you've found the form, you should write down that the integrand equals the form, then multiply through by the denominator.

$$\frac{x+2}{x^2-1}=\frac{A}{x-1}+\frac{B}{x+1}$$

Actually, you're better off writing the denominator on the left-hand side in the factored manner, like this:

$$\frac{x+2}{(x-1)(x+1)}=\frac{A}{x-1}+\frac{B}{x+1}$$

Now multiply through by the denominator $(x-1)(x+1)$ to get

$$x+2=A(x+1)+B(x-1)$$

Anyway, now there are two different ways we can proceed. The first way is to substitute clever values of x. If you put $x=1$, then the $B(x-1)$ term goes away, and you get $A=\frac{3}{2}$. Now if instead you put $x=-1$ in the original equation, the $A(x+1)$ term goes away, so $B=-\frac{1}{2}$.

Alternatively, another way of finding A and B is to take our original equation $x+2=A(x+1)+B(x-1)$ and rewrite it as

$$x+2=(A+B)x+(A-B)$$

Now we can equate coefficients of x to see that $1=A+B$ and $2=A+B$.

You might have noticed that in both of the ways we found A and B, we needed two facts. For the substitution method, we put $x=1$ and then $x=-1$, whereas for the method of equating coefficients, we equated the coefficients of x and also the constant coefficients. We actually could have used one instance of each method (Page 401).

All that's left is to rewrite your integrand as equal to the form again, but this time with the constants filled in. So in our example,

$$\frac{x+2}{x^2-1}=\frac{A}{x-1}+\frac{B}{x+1}=\frac{3/2}{x-1}+\frac{-1/2}{x+1}$$

Now integrate both sides, pulling out the constant factors as you split up the integral:

$$\int\frac{x+2}{x^2-1}dx=\frac{3}{2}\int\frac{1}{x-1}dx-\frac{1}{2}\int\frac{1}{x+1}dx$$

We have successfully busted up our original integral into two integrals which are much simpler

So far, we've seen that we do a long division unless the degree of the top is less than the degree of the bottom; then we factor the denominator; then we write down the form; then we use one of two methods to find the unknown constants. Finally, we write down the integrals of the various pieces.

● Integrating the pieces

We need to see how to integrate the various pieces which remain after you break up the original integral. The simplest type of integral is of the form

$$\int\frac{1}{ax+b}dx$$

To do this, just substitute $t=ax+b$.

The same trick works for a power of a linear factor in the denominator; for example, to find

$$\int\frac{1}{(4x+5)^2}dx$$

Substitute $t = 4x + 5$.

The difficult case involves a quadratic in the bottom, like this:

$$\int \frac{Ax + B}{ax^2 + bx + c} dx$$

Beware! If the quadratic can be factored, then you need to do this first. Then the left is that the quadratic on the bottom cannot be factored. That is, its discriminant $b^2 - 4ac$ is negative (Page 402).

● The method and a big example

Here's the complete method for finding the integral of a rational function:

Step 1-check degrees, divide if necessary: check to see if the degree of the numerator is less than the degree of the denominator. If it is, then you're golden-go on to step 2. If not, do a long division, then proceed to step 2

Step 2-factor the denominator: use the quadratic formula, or guess roots and divide, to factor the denominator of your integrand

Step 3-the form: write down the "form," with undetermined constants, as described on the book (Page 399). Write down an equation like

$$\text{integrand} = \text{form}$$

Step 4-evaluate constants: multiply both sides of this equation by the denominator, then find the constants by (a) substituting clever values of x; (b) equating coefficients; or some combination of (a) and (b). Now you can express your integral as the sum of rational functions which either have constants on the top and powers of linear functions on the bottom, or look like a linear function divided by a quadratic function

Step 5-integrate terms with linear powers on the bottom: solve any integrals whose denominators are powers of linear functions; the answers will involve logs or negative powers of the linear term

Step 6-integrate terms with quadratics on the bottom: for each integral with a nonfactorable quadratic term in the denominator, complete the square, make a change of variables, then possibly split up into two integrals. The first one will involve logs and the second should involve $\tan^{-1}$. If there's only one integral, it could involve either logs or $\tan^{-1}$. This formula is very useful most of the time:

$$\boxed{\int \frac{1}{t^2 + a^2} dt = \frac{1}{a} \tan^{-1}\left(\frac{t}{a}\right) - C}$$

Remember, you don't always need to use all six steps. Sometimes you can go directly to the last step, such as in our example

$$\int \frac{x + 8}{x^2 + 6x + 13} dx$$

CHAPTER 19 Techniques of Integration, Part Two

In this chapter, we'll finish gathering our techniques of integration by taking an extensive look at integrals involving trig functions. Sometimes one has to use trig identities to solve these types of problems; on other occasions there are no trig functions present, so you have to introduce some by making a trig substitution.

19.1 Integrals Involving Trig Identities

There are three families of trig identities which are particularly useful in evaluating integrals. The first family arises from the double-angle formula for $\cos(2x)$. For use in integration, it turns out that the best way to use the formulas is to solve the relevant equation for $\cos^2(x)$ or $\sin^2(x)$. So, we have

$$\boxed{\cos^2(x) = \frac{1}{2}(1+\cos(2x))} \quad \text{and} \quad \boxed{\sin^2(x) = \frac{1}{2}(1-\cos(2x))}$$

It is well worth remembering these identities! In particular, if you ever have to take a square root of $1+\cos(\text{anything})$ or $1-\cos(\text{anything})$, these identities save the day. For example,

$$\int_0^{\pi/2} \sqrt{1-\cos(2x)}\, dx$$

looks pretty nasty, but in fact

$$\int_0^{\pi/2} \sqrt{1-\cos(2x)}\, dx = \int_0^{\pi/2} \sqrt{2\sin^2(x)}\, dx$$

$$= \sqrt{2}\int_0^{\pi/2} |\sin(x)| dx$$

Luckily, when x is between 0 and $\pi/2$, the values of $\sin(x)$ are always greater than or equal to zero, so we have reduced things to

$$\sqrt{2}\int_0^{\pi/2} \sin(x)\, dx$$

Let's move on to the second family of trig identities. These are the Pythagorean identities:

$$\boxed{\sin^2(x)+\cos^2(x)=1} \quad \boxed{\tan^2(x)+1=\sec^2(x)} \quad \boxed{1+\cot^2(x)=\csc^2(x)}$$

These identities are valid for any x, and sometimes they are obviously helpful. For example,

$$\int_0^{\pi} \sqrt{1-\cos^2(x)}\, dx$$

should just be written as

$$\int_0^{\pi} \sqrt{\sin^2(x)}\, dx = \int_0^{\pi} |\sin(x)| dx = \int_0^{\pi} \sin(x)\, dx = 2$$

Now, sometimes you have to apply a devious trick in order to use the above identities. If you see $1+\mathrm{trig}(x)$ or $1-\mathrm{trig}(x)$, where "trig" is some trig function (specifically sine, cosine, secant, or cosecant), in the denominator of an integral, consider multiplying by the conjugate expression. For example, to find

$$\int \frac{1}{\sec(x)-1} dx$$

multiply top and bottom by the conjugate expression of the denominator, which in this case is $\sec(x)+1$. That is

$$\int \frac{1}{\sec(x)-1}dx = \int \frac{1}{\sec(x)-1} \times \frac{\sec(x)+1}{\sec(x)+1}dx$$

$$= \int \frac{\sec(x)+1}{\sec^2(x)-1}dx = \int \frac{\sec(x)+1}{\tan^2(x)}dx = \int \frac{\sec(x)}{\tan^2(x)}dx + \int \frac{1}{\tan^2(x)}dx$$

$$= \int \frac{1/\cos(x)}{\sin^2(x)/\cos^2(x)}dx + \int \cot^2(x)\,dx = \int \frac{\cos(x)}{\sin^2(x)}dx + \int (\csc^2(x)-1)dx$$

$$= -\csc(x) - \cot(x) - x + C$$

Let's look at the third family of identities, the so-called products-to-sums identities:

$$\boxed{\begin{aligned} \cos(A)\cos(B) &= \frac{1}{2}(\cos(A-B)+\cos(A+B)) \\ \sin(A)\sin(B) &= \frac{1}{2}(\cos(A-B)-\cos(A+B)) \\ \sin(A)\cos(B) &= \frac{1}{2}(\sin(A-B)+\sin(A+B)) \end{aligned}}$$

It's quite a pain in the butt to remember these. Actually, they all follow from the expressions for $\cos(A \pm B)$ and $\sin(A \pm B)$, so if you have those down, you can reverse engineer the above identities from them. These identities are quite indispensable for finding integrals like

$$\int \cos(3x)\sin(19x)\,dx$$

$$\int \cos(3x)\sin(19x)\,dx = \frac{1}{2}\int (\sin(19x-3x)+\sin(19x+3x))dx$$

$$= \frac{1}{2}\int (\sin(16x)+\sin(22x))dx = \frac{1}{2}\left(-\frac{\cos(16x)}{16} - \frac{\cos(22x)}{22}\right) + C = -\frac{\cos(16x)}{32} - \frac{\cos(22x)}{44} + C$$

19.2 Integrals Involving Powers of Trig Functions

Now we'll see how to find certain integrals which have powers of trig functions in their integrands. For example, how would you find $\int \cos^7(x)\sin^{10}(x)\,dx$ or $\int \sec^6(x)\,dx$? Unfortunately, these types of integrals require different techniques, depending on which trig function or functions you're dealing with.

- Powers of sin and/or cos

Here's the golden rule: if one of the powers of $\sin(x)$ or $\cos(x)$ is odd, then grab it and don't let it get away-it is your friend! (If they are both odd, then take the one with the lowest power as your friend.) If you've grabbed your odd power, then you need to pull out one power to go with the dx; then deal with what's left (which is now an even power) by using one of the identities

$$\boxed{\cos^2(x) = 1-\sin^2(x)} \quad \text{or} \quad \boxed{\sin^2(x) = 1-\cos^2(x)}$$

Anyway, the best way to see how the technique of pulling out one power from the odd power works is by looking at an example (Page 413)

$$\int \cos^7(x)\sin^{10}(x)\,dx = \int \cos^6(x)\sin^{10}(x)\cos(x)\,dx$$

$$= \int (1-\sin^2(x))^3\sin^{10}(x)\cos(x)\,dx$$

Now if we put $t=\sin(x)$, then $dt=\cos(x)\,dx$, so it's easy to get this integral over to t-land-it's just

$$\int(1-t^2)^3t^{10}dt=\int(1-3t^2+3t^4-t^6)t^{10}dt=\int(t^{10}-3t^{12}+3t^{14}-t^{16})dt$$

which works out to be

$$\frac{t^{11}}{11}-\frac{3t^{13}}{13}+\frac{t^{15}}{5}-\frac{t^{17}}{17}+C$$

Now, what if neither power is odd? Well, if both powers are even-for example, if you had to work out $\int\cos^2(x)\sin^4(x)\,dx$-you should use the double-angle formulas. We just saw them in the previous section, but here they are again for reference:

$$\boxed{\cos^2(x)=\frac{1}{2}(1+\cos(2x))} \quad \text{and} \quad \boxed{\sin^2(x)=\frac{1}{2}(1-\cos(2x))}$$

Now you can just replace everything in sight, and you'll get a whole bunch of simpler integrals which are various powers of cosines. You then need to find them using the same techniques as we have just used, depending on whether the power in each integral is even or odd. In our example

$$\int\cos^2(x)\sin^4(x)\,dx=\int\frac{1}{2}(1+\cos(2x))\left(\frac{1}{2}(1-\cos(2x))\right)^2dx$$

$$=\frac{1}{8}\int(1-\cos(2x)-\cos^2(2x)+\cos^3(2x))dx$$

$$=\frac{1}{8}\int\left(1-\cos(2x)-\frac{1}{2}(1+\cos(4x))+(1-\sin^2(2x))\cos(2x)\right)dx$$

$$=\frac{1}{8}\left(x-\frac{\sin(2x)}{2}-\frac{x}{2}-\frac{\sin(4x)}{8}+\frac{\sin(2x)}{2}-\frac{\sin^3(2x)}{6}\right)+C$$

$$=\frac{x}{16}-\frac{\sin(4x)}{64}-\frac{\sin^3(2x)}{48}+C$$

- Powers of tan

Consider $\int\tan^n(x)\,dx$, where n is some integer. Let's look at the first couple of cases. For $n=1$, we need to know how to do $\int\tan(x)\,dx$. This is a pretty standard integral, which you can solve by setting $t=\cos(x)$, noting that $dt=-\sin(x)\,dx$:

$$\int\tan(x)\,dx=\int\frac{\sin(x)}{\cos(x)}dx=-\int\frac{dt}{t}=-\ln|t|+C=-\ln|\cos(x)|+C$$

The answer can also be written as $\ln|\sec(x)|+C$.

How about $n=2$? For this case, and indeed other cases, it's essential to use the Pythagorean identity

$$\boxed{\tan^2(x)=\sec^2(x)-1}$$

which we looked at in the previous section. So we have

$$\int\tan^2(x)\,dx=\int(\sec^2(x)-1)dx=\tan(x)-x+C$$

To do higher powers $(n\geq 3)$, you have to extract $\tan^2(x)$ and change it into $(\sec^2(x)-1)$. This gives you two integrals. The first can be done by substituting $t=\tan(x)$ and using $dt=$

$\sec^2(x)\,dx$. The second is a lower power of $\tan(x)$ and you can just repeat the method. For example, how would you find $\int \tan^6(x)\,dx$? Let's see: (Page 415)

$$\int \tan^6(x)\,dx = \int \tan^4(x)\tan^2(x)\,dx = \int \tan^4(x)\,(\sec^2(x) - 1)dx$$
$$= \int \tan^4(x)\sec^2(x)\,dx - \int \tan^4(x)\,dx$$

● Powers of sec

Yup, this one really sucks, except for $\int \sec^2(x)\,dx$, which is easy. Let's start with the first power, $\int \sec(x)\,dx$. There are many ways of finding this integral. The easiest involves a cool trick that is well worth remembering, as it's a real timesaver. Unfortunately it's the sort of trick that is completely counterintuitive, and it boggles the mind that anyone even thought of it in the first place. The idea is to multiply top and bottom by the bizarre quantity $(\sec(x) + \tan(x))$. Watch and be amazed:

$$\int \sec(x)\,dx = \int \sec(x) \times \frac{\sec(x) + \tan(x)}{\sec(x) + \tan(x)}\,dx = \int \frac{\sec^2(x) + \sec(x)\tan(x)}{\sec(x) + \tan(x)}\,dx$$
$$= \ln|\sec(x) + \tan(x)| + C$$

since the derivative of the denominator $\sec(x) + \tan(x)$ is miraculously equal to the numerator

How about the second power of $\sec(x)$? Not much to this one:

$$\int \sec^2(x)\,dx = \tan(x) + C$$

That was easy. Unfortunately, it gets pretty messy for larger powers. The standard idea is to pull out $\sec^2(x)$ (which is similar to what we did with powers of $\tan(x)$) and integrate by parts, using $dv = \sec^2(x)\,dx$ and u as the rest of the powers of $\sec(x)$. This means that $v = \tan(x)$ (remember, we don't need a constant here). When you do the integration by parts, you will of course get a new integral; the integrand should be a lower power of $\sec(x)$ multiplied by $\tan^2(x)$. Once again, we have to use $\tan^2(x) = \sec^2(x) - 1$ and get two integrals. One of them is a multiple of the original integral! You have to put this back on the left-hand side. The other one is a lower power of $\sec(x)$, and you have to repeat the whole process until you get down to $\int \sec(x)\,dx$ or $\int \sec^2(x)\,dx$, both of which we now know how to do.

A formidable example: $\int \sec^6(x)\,dx$ (Page 417)

● Powers of cot

These work just like powers of $\tan(x)$. You pull out $\cot^2(x)$ and use the Pythagorean identity

$$\boxed{\cot^2(x) = \csc^2(x) - 1}$$

Just beware that when you set $t = \cot(x)$, you have $dt = -\csc^2(x)\,dx$.

● Powers of csc

These work just like powers of $\sec(x)$. You pull out $\csc^2(x)$ and integrate by parts, using $dv = \csc^2(x)\,dx$. Beware: you now have $v = -\cot(x)$, and du also involves a minus sign which you have to worry about.

● Reduction formulas

The methods of the last four sections all involve knocking the power of the trig function you're dealing with down by 2, then repeating the process. Let's try to write out the method in general. First, we're dealing with $\int \tan^n(x)\,dx$, so we'll give it a name: I_n (for integral number n). That is,

$$I_n = \int \tan^n(x)\,dx$$

(Page 419)

$$I_n = \frac{1}{n-1}\tan^{n-1}(x) - I_{n-2}$$

The above equation is called a *reduction formula*, since it helps us reduce the number n to a smaller number $n-2$.

The method also works for definite integrals. For example, how would you find the definite integral $\int_0^{\pi/2}\cos^8(x)\,dx$? You could use the double-angle formulas, but that would be a pain in the ass. (Try it if you don't believe me!) Instead, let's set (Page 420)

$$I_n = \int_0^{\pi/2}\cos^n(x)\,dx$$

$$I_n = \frac{n-1}{n}I_{n-2}$$

By the way, reduction formulas don't have to involve trig functions. For example, if (Page 421)

$$I_n = \int x^n e^x dx$$

$$I_n = x^n e^x - nI_{n-1}$$

19.3 Integrals Involving Trig Substitutions

Now let's look at how to do integrals involving an odd power of the square root of a quadratic. Here are some examples of the type of integral we're considering:

$$\int \frac{dx}{x^3\sqrt{x^2-4}} \quad \text{or} \quad \int \frac{x^2}{(9-x^2)^{3/2}}dx \quad \text{or} \quad \int (x^2+15)^{-5/2}dx$$

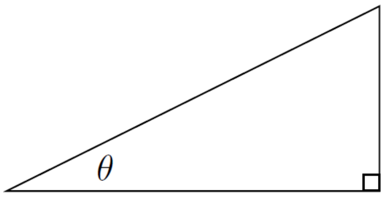
The basic idea is that there are three types, corresponding to whether you have to worry about a^2-x^2, x^2+a^2, or x^2-a^2. Here a is just some number. Each of these three types requires a different substitution. Most of the time, after substituting, you end up with an integral involving powers of trig functions, which is where the previous section comes in.

● Type 1: $\sqrt{a^2-x^2}$

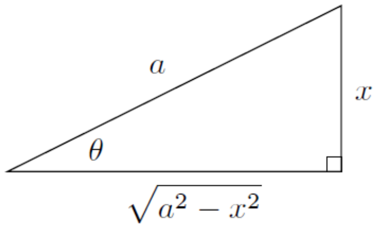
If you have an integral involving an odd power of $\sqrt{a^2-x^2}$, the correct substitution to use is $x = a\sin(\theta)$. (You could use $x = a\cos(\theta)$ if you prefer, but there would be no advantage to it, so stick with sine.) The reason that this substitution is effective is that

$$a^2 - x^2 = a^2 - a^2\sin^2(\theta) = a^2(1-\sin^2(\theta)) = a^2\cos^2(\theta)$$

and now you can easily take a square root. Remember that if you are changing variables from x to θ, you have to go from x-land to θ-land. That is, everything about the integral has to be in terms of θ, not x. In particular, we'll need to replace dx by something in θ and $d\theta$. No problem-just differentiate the equation $x = a\sin(\theta)$ to get $dx = a\cos(\theta)\,d\theta$. Anyway, now we can hopefully do the integral in θ-land, but in the end we have to change the answer back to x-land. To do this, it will be useful to draw the following right-angled triangle with one angle equal to θ:



Now we know $\sin(\theta) = x/a$, so



Let's see how it works in practice (Page 422)

Before we go on to Type 2, do you see that we've been a little careless here? (Page 423)

$$(\cos^2(\theta))^{3/2} = \left(\sqrt{\cos^2(\theta)}\right)^3 = |\cos^3(\theta)|$$

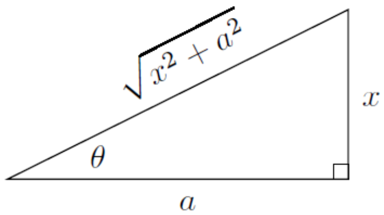
Luckily, the absolute value signs turn out to be unnecessary for Type 1 and also for Type 2 below (but not for Type 3), so we were right all along.

● Type 2: $\sqrt{x^2 + a^2}$

If an integral involves an odd power of $\sqrt{x^2 + a^2}$, the correct substitution is $x = a\tan(\theta)$. This works because

$$x^2 + a^2 = a^2\tan^2(\theta) + a^2 = a^2(\tan^2(\theta) + 1) = a^2\sec^2(\theta)$$

Also, we'll need to know that $dx = a\sec^2(\theta)\,d\theta$. Since $\tan(\theta) = x/a$, the triangle now looks like this:

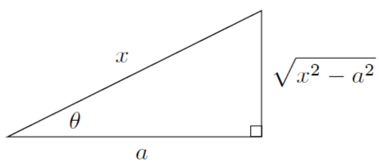


● Type 3: $\sqrt{x^2 - a^2}$

Finally, how about integrals involving an odd power of $\sqrt{x^2 - a^2}$? Now the correct substitution is $x = a\sec(\theta)$, since

$$x^2 - a^2 = a^2\sec^2(\theta) - a^2 = a^2(\sec^2(\theta) - 1) = a^2\tan^2(\theta)$$

and you can easily take square roots. To make the substitution, we'll also need the fact that $dx = a\sec(\theta)\tan(\theta)\,d\theta$. Since $\sec(\theta) = x/a$, the triangle looks like this:



For example, to find

$$\int \frac{dx}{x^3\sqrt{x^2 - 4}}$$

(Page 425) Actually, this time it's wrong to replace $\sqrt{4\tan^2(\theta)}$ by $2\tan(\theta)$; this is only correct if $x > 0$ in the original integral.

● Completing the square and trig substitutions

Now, one other important point before we summarize the situation. From time to time, you might want to solve an integral involving an odd power of $\sqrt{\pm x^2 + ax + b}$. That is, you now have a linear term ax to complicate matters. The technique is simple: complete the square first and substitute to get it into one of the three types that we've investigated. For example, to evaluate

$$\int (x^2 - 4x + 19)^{-5/2} dx$$

first complete the square

$$\int ((x-2)^2 + 15)^{-5/2} dx$$

Now let $t = x - 2$, so $dt = dx$, and in t-land the integral becomes

$$\int (t^2 + 15)^{-5/2} dt$$

$$\int (t^2 + 15)^{-5/2} dt = \frac{1}{225}\left(\frac{t}{\sqrt{t^2+15}} - \frac{t^3}{3(t^2+15)^{3/2}}\right) + C$$

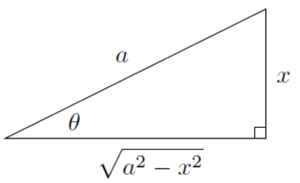
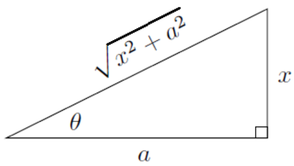
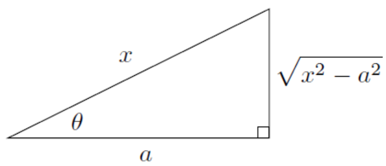
so replacing t now by $x - 2$, we see that

$$\int (x^2 - 4x + 19)^{-5/2} dx = \frac{1}{225}\left(\frac{x-2}{\sqrt{x^2-4x+19}} - \frac{(x-2)^3}{3(x^2-4x+19)^{3/2}}\right) + C$$

The moral of the story, both here and when using partial fractions, is that a quadratic with a linear term can be made into a quadratic without one by completing the square and substituting.

● Summary of trig substitutions

To summarize the three main types we've looked at, here's a table that shows the appropriate substitutions and triangles for each type:

Type 1: $\sqrt{a^2 - x^2}$	Type 2: $\sqrt{x^2 + a^2}$	Type 3: $\sqrt{x^2 - a^2}$
Set $x = a\sin(\theta)$	Set $x = a\tan(\theta)$	Set $x = a\sec(\theta)$
$dx = a\cos(\theta)\,d\theta$	$dx = a\sec^2(\theta)\,d\theta$	$dx = a\sec(\theta)\tan(\theta)\,d\theta$
$a^2 - x^2 = a^2\cos^2(\theta)$	$x^2 + a^2 = a^2\sec^2(\theta)$	$x^2 - a^2 = a^2\tan^2(\theta)$
		

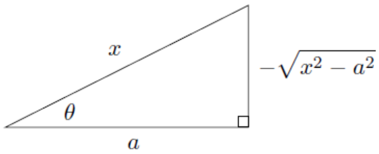
The next section discusses the technical point about when (and why) you can drop the absolute value signs when you take square roots of quantities.

● Technicalities of square roots and trig substitutions

Now, think back to Type 1 above. We know that the range of $\sin^{-1}$ is $[-\pi/2,\ \pi/2]$; this means that θ is in the first or fourth quadrant, so $\cos(\theta)$. We don't need any absolute values!

The same goes for Type 2. In that case, we'd really like to simplify $\sqrt{a^2\sec^2(\theta)}$ as $a\sec(\theta)$. We have $x = a\tan(\theta)$, so $\theta = \tan^{-1}(x/a)$. The range of $\tan^{-1}$ is $(-\pi/2,\ \pi/2)$, so θ is once again in the first or fourth quadrant. This means that $\sec(\theta)$ is always positive, so again, we don't need absolute values.

Everything goes wrong in Type 3, unfortunately. Here we need to deal with $\sqrt{a^2\tan^2(\theta)}$, but this isn't always equal to $a\tan(\theta)$. You see, since $x = a\sec(\theta)$, we have $\theta = \sec^{-1}(x/a)$. You'll see that the range of $\sec^{-1}$ is the interval $[0, \pi]$, except for the point $\pi/2$. So θ is in the first or second quadrant, and $\tan(\theta)$ could be positive or negative. At least it has the same sign as x does, as you can see by looking at the graph of $y = \sec^{-1}(x)$. So, if $x < 0$, you have to write $-a\tan(\theta)$ instead. In that case, the triangle actually looks like this: (Page 427)



19.4 Overview of Techniques of Integration

We've now built up quite a toolkit of techniques of integration. Now the question is, given an integral, which technique do you use?

Here are some general guidelines to help you out:

● If an "obvious" substitution comes to mind, try it. For example, if one factor of the integrand is the derivative of another piece of the integrand, try substituting t for that other piece.

● If something like $\sqrt[n]{ax+b}$ appears in the integrand, try substituting $t = \sqrt[n]{ax+b}$.

● To integrate a rational function (that is, a quotient of polynomials), see if the top is a multiple of the derivative of the bottom. If so, you can just substitute $t = \text{denominator}$. Otherwise, use partial fractions.

● After checking that no obvious substitution looks as if it will work, use the techniques from the beginning of this chapter to find integrals involving:

-functions containing $\sqrt{1+\cos(x)}$ or $\sqrt{1-\cos(x)}$: in this case, use the double-angle formula;

-functions involving one of $1-\sin^2(x)$, $1-\cos^2(x)$, $1+\tan^2(x)$, $\sec^2(x)-1$, $\csc^2(x)-1$, or $1+\cot^2(x)$: in this case, use one of the Pythagorean identities $\sin^2(x)+\cos^2(x)=1$, $\tan^2(x)+1=\sec^2(x)$, or $1+\cot^2(x)=\csc^2(x)$;

-functions with $1 \pm \sin(x)$ (or similar) in the denominator: in this case, multiply and divide by the conjugate expression and try to use the Pythagorean identities;

-functions containing products like $\cos(mx)\cos(nx)$, $\sin(mx)\sin(nx)$, or $\sin(mx)\cos(nx)$: in this case, use the products-to-sums identities; or

-powers of trig functions: you'll just have to learn the individual techniques in Sections 19.2.1 through 19.2.5 above.

● If the integrand involves $\sqrt{x^2-a^2}$ or any odd power of this (for example $(x^2-a^2)^{3/2}$, $(x^2-a^2)^{5/2}$, and so on), or $\sqrt{x^2+a^2}$ or $\sqrt{a^2-x^2}$ or an odd power of any of these last two, then use a trig substitution (after checking that there's no obvious substitution). If the quadratic includes a linear term, complete the square first.

● If the integrand is a product and no obvious substitution comes to mind, try integration by parts.

● If no substitution appeals, then a good rule of thumb is that functions involving a power of $\ln(x)$ or an inverse trig function should be integrated by parts. In that case, let u be the power of $\ln(x)$ or the inverse trig function as appropriate.

CHAPTER 20 Improper Integrals: Basic Concepts

20.1 Convergence and Divergence

What is an improper integral, anyway? In Chapter 16, we saw that the integral

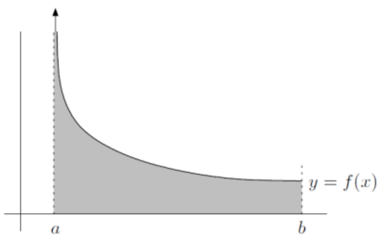
$$\int_a^b f(x)dx$$

certainly makes sense if f is a bounded function on $[a,b]$ which is continuous except at a finite number of places. If f has infinitely many discontinuities, the integral might still make sense, or it might be totally screwed up. What if f isn't bounded? This means that the values of $f(x)$ manage to get really large (positively or negatively or both) while x is in the interval $[a,b]$. This sort of thing typically happens when f has a vertical asymptote somewhere in this interval: the function blows up there and can't be bounded. This causes the above integral to be improper.

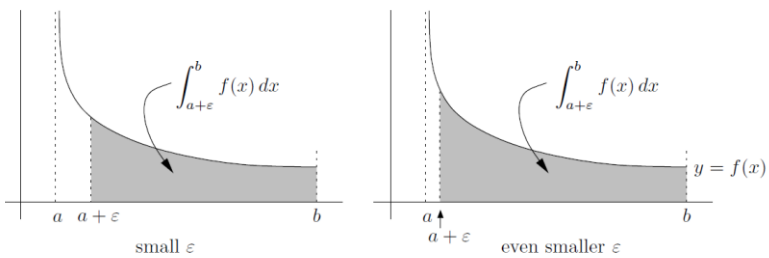
The integral $\int_a^b f(x)dx$ is ***improper*** if any of the following conditions apply:

1. f isn't bounded in the **closed** interval $[a,b]$;
2. $b=\infty$; or
3. $a=-\infty$

For now, let's concentrate on what happens if the first of these conditions fails. There is a simple case of when our function f has a vertical asymptote at $x=a$. The situation looks something like this:



Since the region never stops going up, surely its area should be infinite, right? Not necessarily. A mathematical miracle can occur if the region is skinny enough, and the area can actually be finite. To see how a region can be unbounded yet have a finite area, we'll use limits once again. Here's the idea: let ε be a small positive number; then you can integrate f over the region $[a+\varepsilon,b]$, since f is bounded there. You'll get some nice finite number. Now, replay the situation but with an even smaller ε. You get a new finite number. The situation now looks something like this:



The smaller ε is, the closer our (bounded) approximating region is to the actual unbounded region. This suggests that we should continue the process with smaller and smaller ε, and see if the numbers we get have a limit L as $\varepsilon\to0^+$. If so, then we interpret L square units to be the value of the area we're looking for. In that case, we say that the integral $\int_a^b f(x)dx$ ***converges*** to L. If there's no limit, then we can't find a meaningful answer for the area, so we give up and say that the above integral ***diverges***. Note that **if the integral isn't improper, it automatically converges**!

Now, here's a summary of the situation when you have a blow-up point at $x = a$:

> if $f(x)$ is unbounded for x near a only, then set
> $$\int_a^b f(x)dx = \lim_{\varepsilon\to 0^+} \int_{a+\varepsilon}^b f(x)dx$$

provided that the limit exists. If it does, then the integral converges; if not, the integral diverges. Just like any limit, the above one may fail to exist because it might be ∞ or $-\infty$, or things might oscillate around too much as ε tends to 0^+.

This brings us to an important point. When we look at an improper integral, the most important thing we need to find out is whether it converges or diverges. It's much less important to know what the integral converges to (assuming it converges).

● Some examples of improper integrals

Consider the integrals

$$\int_0^1 \frac{1}{x}dx \quad \text{and} \quad \int_0^1 \frac{1}{\sqrt{x}}dx$$

In the first case, we have

$$\int_0^1 \frac{1}{x}dx = \lim_{\varepsilon\to 0^+} \int_\varepsilon^1 \frac{1}{x}dx = \lim_{\varepsilon\to 0^+} \ln|x| \Big|_\varepsilon^1 = \lim_{\varepsilon\to 0^+} (\ln(1) - \ln(\varepsilon)) = \infty$$

Since we got ∞, the improper integral $\int_0^1 1/xdx$ must diverge. How about the other integral? Using the formula again, we have

$$\int_0^1 \frac{1}{\sqrt{x}}dx = \lim_{\varepsilon\to 0^+} \int_\varepsilon^1 \frac{1}{x^{1/2}}dx = \lim_{\varepsilon\to 0^+} 2x^{1/2} \Big|_\varepsilon^1 = \lim_{\varepsilon\to 0^+} \left(2\sqrt{1} - 2\sqrt{\varepsilon}\right) = 2$$

Now, here's a really important point. Suppose you have an improper integral $\int_a^b f(x)dx$, where f has a vertical asymptote at $x = a$ only, and you want to know if the integral converges or diverges. Then the value of b doesn't matter! You can change it to any finite number bigger than a, so long as you don't pick up any new vertical asymptotes or blow-up points.

● Other blow-up points

In the integral $\int_a^b f(x)dx$, if f has only one blow-up point at the right-hand limit of integration b (instead of a), then we can play the same game as we did above. The only difference is that this time we have to approach b from the left instead of the right. So

> if $f(x)$ is unbounded for x near b only, then set
> $$\int_a^b f(x)dx = \lim_{\varepsilon\to 0^+} \int_a^{b-\varepsilon} f(x)dx$$

if the limit exists; if it doesn't exist, then as before, the integral diverges.

But what if f has a blow-up point at some number c in the interior of the interval? In this case, if f is bounded everywhere on $[a, b]$ except near some point c in the interior (a, b), we have to split the integral into the two pieces

$$\int_a^c f(x)dx \quad \text{and} \quad \int_c^b f(x)dx$$

We can see that the above integrals are

$$\lim_{\varepsilon\to 0^+}\int_a^{c-\varepsilon} f(x)dx \quad \text{and} \quad \lim_{\varepsilon\to 0^+}\int_{c+\varepsilon}^{b} f(x)dx$$

respectively. Here's the essential point: the whole integral $\int_a^b f(x)dx$ only converges if both pieces above converge. If either piece diverges, so does the whole thing.

This example inspires our first main technique: to investigate an improper integral, split it up into pieces, if necessary. Each piece has to have at most one problem spot, which must be at one of the limits of integration. (For the moment, the term "problem spot" means the same thing as "blow-up point," but in the next section we'll see a different sort of problem spot that isn't a blow-up point.)

For example, to analyze the integral (Page 436)

$$I = \int_0^3 \frac{1}{x(x-1)(x+1)(x-2)}dx$$

20.2 Integrals over Unbounded Regions

Now, we still have to look at what happens when one or both of the limits of integration are infinite; this means that the region of integration is unbounded. To handle

$$\int_a^\infty f(x)dx$$

where a is any finite number and f has no blow-up points in $[a, \infty)$, let's use another limiting technique. This time, we integrate over the region $[a, N]$, where N is a massively large number. This will give us a nice finite value. Then repeat but with an even larger N to get a new value. Continue onward and see what happens to the values of the integrals. If they have a limit, then the integral converges. Otherwise, it diverges. In symbols, we are defining

$$\boxed{\int_a^\infty f(x)dx = \lim_{N\to\infty}\int_a^N f(x)dx}$$

provided that the limit exists; in this case, the integral converges. Otherwise, it diverges. The value of a is irrelevant. So long as you don't pick up any new blowup points of f, the value of a doesn't affect whether the improper integral converges or diverges. The only thing that really matters is how $f(x)$ behaves when x is very large indeed.

In a similar manner to the above definition, if f has no blow-up points in $(-\infty, b]$, then

$$\boxed{\int_{-\infty}^b f(x)dx = \lim_{N\to\infty}\int_{-N}^b f(x)dx}$$

What if f has no blow-up points anywhere and we want to find

$$\int_{-\infty}^\infty f(x)dx$$

Although there are no blow-up points, there are still two problem spots: ∞ and $-\infty$. We have to split the above integral into two pieces so that each one has only one problem spot

$$\int_{-\infty}^{0} f(x)dx \quad \text{and} \quad \int_{0}^{\infty} f(x)dx$$

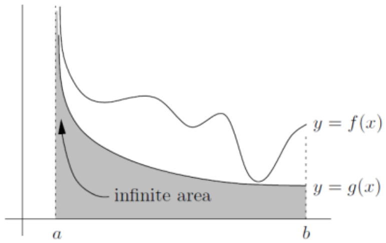
Here are some examples involving an unbounded region of integration (Page 438), and there is a special case of the so-called comparison test.

20.3 The Comparison Test (Theory)

Suppose we have two functions which are never negative, at least in some region of interest. If the first function is bigger than the second function, and the integral of the second function (over our region) diverges, then the integral of the first function (over the same region) also diverges. Mathematically, it looks like this. Let's say we want to know something about $\int_a^b f(x)dx$, but we only know something about $\int_a^b g(x)dx$. If $f(x) \geq g(x) \geq 0$ for x in the interval (a, b), and we know that $\int_a^b g(x)dx$ diverges, then so does $\int_a^b f(x)dx$. In fact, since $f(x) \geq g(x)$, we can write

$$\int_a^b f(x)dx \geq \int_a^b g(x)dx = \infty$$

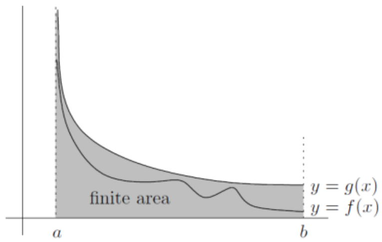
So the first integral also diverges. The situation is even clearer when one looks at a picture:



On the other hand, for convergence, it is the other way around. Here, if we want to know about $\int_a^b f(x)dx$ and we know that $\int_a^b g(x)dx$ converges, we'd better hope that $f(x) \leq g(x)$. You might say that we want f to be "controlled" by g. Well, then we'd get convergence (still assuming that both functions are positive). So, if $0 \leq f(x) \leq g(x)$ on (a, b) and $\int_a^b g(x)dx$ converges, then so does $\int_a^b f(x)dx$, Mathematically,

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx < \infty$$

so both integrals converge (noting that the left-hand integral is positive, so it can't diverge down to $-\infty$). The picture looks like this:



Beware: (Page 440)

20.4 The Limit Comparison Test (Theory)

The comparison test uses the improper integral of one function to get information about an improper integral of another function. The limit comparison test does the same thing, except that we don't actually need one function to be bigger than the other. Instead, we need the two functions to be just about the same. Here's the basic idea: suppose that two functions f and g are very close to each other at the blow-up point $x = a$ (and have no other blow-up points). Then $\int_a^b f(x)dx$ and $\int_a^b g(x)dx$ either both diverge or both converge. Their behavior is identical.

● Functions asymptotic to each other

Suppose we have two functions f and g such that

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

This means that when x is near a, the ratio $f(x)/g(x)$ is close to 1. If the ratio were equal to 1, then $f(x)$ would equal $g(x)$. Since the ratio is only close to 1, then $f(x)$ is "very close" to $g(x)$.

So, we'll say that $f(x) \sim g(x)$ as $x \to a$ if the limit of the ratio is 1. That is,

$$\boxed{f(x) \sim g(x) \text{ as } x \to a \text{ means the same thing as } \lim_{x\to a}\frac{f(x)}{g(x)} = 1}$$

All we've done is to rewrite each limit in a different form, but it is a very convenient form. Indeed, you can take powers of asymptotic relations and get new ones. For example, knowing that $\sin(x) \sim x$ as $x \to 0$, we can immediately write that $\sin^3(x) \sim x^3$ as $x \to 0$, or even that $1/\sin(x) \sim 1/x$ as $x \to 0$. You can also replace x by any other quantity that goes to 0 as x does, such as a power of x. For example $\sin(4x^7) \sim 4x^7$ as $x \to 0$. You can even multiply or divide two relations by each other, provided that the limit is at the same value of x for both asymptotic relations. For example, we know that $\tan(x) \sim x$ as $x \to 0$ since $\lim_{x\to 0}\frac{\tan(x)}{x} = 1$. So we can multiply $\tan(x) \sim x$ and $\sin(x) \sim x$ (both as $x \to 0$) together to get the asymptotic relation $\tan(x)\sin(x) \sim x^2$ as $x \to 0$

What you cannot do is add or subtract these relations. For example, if you start with $\tan(x) \sim x$ and $\sin(x) \sim x$ as $x \to 0$, you can't just subtract the second relation from the first to get $\tan(x) - \sin(x) \sim x - x$.

● The statement of the test

If $f(x) \sim g(x)$ as $x \to a$, and neither function has any problem spots anywhere else on the interval $[a, b]$, then the integrals $\int_a^b f(x)dx$ and $\int_a^b g(x)dx$ both diverge or both converge. (If they both converge, then the values they converge to may be different.) This is one case of the *limit comparison test*. Here's a sneak preview of its power; we'll see many more examples in the next chapter. Suppose we want to know whether

$$\int_0^1 \frac{1}{\sin(\sqrt{x})}dx$$

converges or diverges (Page 443)

Of course, there are cases of the test which apply when the blow-up point is at b, or when the region of integration is unbounded.

In particular (Page 444). A quick comment: most textbooks have a different statement of the limit comparison test. In particular, the limit of $f(x)/g(x)$ doesn't actually have to be 1-it could be any positive number and the above argument would still work (after a slight modification).

20.5 The p-test (Theory)

Now that we have the comparison test and limit comparison test, we need to know how to use them. Our basic strategy, which will be greatly elaborated upon in the next chapter, will be to pick a function g which we can compare our function f with. Hopefully g is simple enough that we can at least say whether its integral (over the region under consideration) converges or diverges.

The question is, what are some functions we could choose as g? Well, the most useful are the functions $1/x^p$ for some $p > 0$. The p-test:

- (**p-test**, $\int^{\infty}$ **version**) For any finite $a > 0$, the integral

$$\int_a^{\infty} \frac{1}{x^p} dx$$

converges if $p > 1$ and diverges if $p \leq 1$

- (**p-test**, $\int_0$ **version**) For any finite $a > 0$, the integral

$$\int_0^{a} \frac{1}{x^p} dx$$

converges if $p < 1$ and diverges if $p \geq 1$.

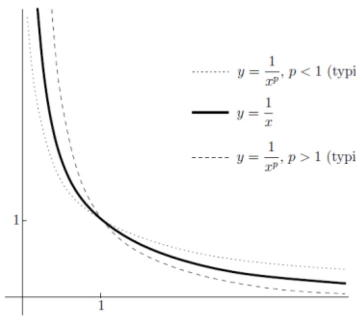
Notice that the two versions of the test are basically opposites: except for when $p = 1$, one of the integrals

$$\int_0^{a} \frac{1}{x^p} dx \quad \text{and} \quad \int_a^{\infty} \frac{1}{x^p} dx$$

converges and the other one diverges. The case $p = 1$ corresponds to $1/x$, and as we already know, both of the integrals diverge in this case.

One way to remember the correct version of the test is to remember what happens with $1/x^2$ and $1/\sqrt{x}$. I just remember the two little facts:

$$\int_a^{\infty} \frac{1}{x^2} dx \text{ converges, and so does } \int_0^{a} \frac{1}{\sqrt{x}} dx$$



20.6 The Absolute Convergence Test

One of the assumptions in the comparison test is that the functions f and g are always nonnegative. What if you want to investigate the behavior of a function which is sometimes negative? Well, if the function is always negative, you could just pull out a minus sign and reduce it to the case of a positive function. We'll see an example of this in the next chapter. On the other hand, if the function keeps oscillating between positive and negative values throughout the region of integration, you can appeal to the **absolute convergence test**. Here's what it says:

If $\int_a^b |f(x)|dx$ converges, then so does $\int_a^b f(x)dx$

This also works on infinite regions of integration (such as $[a, \infty)$ instead of $[a, b]$). Watch out: if the absolute-value version of the original integral diverges, then the original integral could still converge!

Why is the above test useful? Well, for one thing, $|f(x)|$ is always nonnegative, so you can use the comparison test on improper integrals involving it. For example, consider the improper integral (Page 447)

$$\int_1^\infty \frac{\sin(x)}{x^2} dx$$

CHAPTER 21 Improper Integrals: How to Solve Problems

Let's get practical and look at a lot of examples of improper integrals.

21.1 How to Get Started

Our first task is to split up the integral as appropriate, and our second task will be to deal with what happens if f is sometimes negative.

● Splitting up the integral

Here's the basic plan of attack:

1. **Identify all the problem spots** in the region $[a, b]$
2. **Split up the integral** into enough pieces so that each new piece has at most one problem spot, which occurs at one of the endpoints of the integral
3. **Look at each piece individually. If any one piece diverges, so does the whole thing**. The only way the original improper integral can converge is if each piece converges.

The key point is that all the pieces have to converge in order for the whole integral to converge

Here's an important case: what if there are no problem spots? That is, (Page 452). In summary, **if there are no problem spots, the integral automatically converges!**

● How to deal with negative function values

If $f(x)$ takes on negative values for some x in $[a, b]$, which often happens when trig functions or logs are present, you need to take special care. Luckily you can often reduce matters to integrals with only positive integrands. Here are three ways to deal with negative function values:

1. If the integrand $f(x)$ is both positive and negative as x ranges over $[a, b]$, you should consider trying the *absolute convergence test*. As we saw in Section 20.6 of the previous chapter, this says that

If $\int_a^b \lvert f(x)\rvert dx$ converges, then so does $\int_a^b f(x)dx$

In general, don't forget this important point: the absolute convergence test only helps you show that an integral converges. That is, **you cannot use the absolute convergence test to show that an integral diverges!**

2. Suppose that the integrand $f(x)$ is **always** negative (or zero) on $[a, b]$. That is, $f(x) \leq 0$ on $[a, b]$. If this is true, you can write

$$\int_a^b f(x)dx = -\int_a^b (-f(x))dx$$

So what? Well, $-f(x)$ is now always nonnegative, so you can use the comparison test or the p-test to see whether $\int_a^b (-f(x))dx$ converges or diverges. Of course, if this integral converges, so does $\int_a^b f(x)dx$, and similarly if $\int_a^b (-f(x))dx$ diverges, so does $\int_a^b f(x)dx$

3. If neither of the previous two cases seems to apply, you may be able to use the formal definition of the improper integral to see what's going on. An example of this is

$$\int_0^\infty \cos(x)\, dx$$

which we looked at on page 448.

This is not the end of the story. There are slightly freaky improper integrals which converge, but which are not absolutely convergent. These sorts of improper integrals seem to come up quite often in actual physics and engineering applications, but they are beyond the scope of this book. So, it's time to go back and review the integral tests.

21.2 Summary of Integral Tests

The most valuable tools you have at your disposal are the comparison test, the limit comparison test, and the p-test. We looked at these tests from a theoretical point of view in the previous chapter; here are the statements once again, for reference. **In all the tests below, the integrand $f(x)$ is assumed to be positive on the region of integration**

● **Comparison test, divergence version:** if you think that $\int_a^b f(x)dx$ diverges, find a smaller function whose integral also diverges. That is, find a nonnegative function g such that $f(x) \geq g(x)$ on (a,b), and such that $\int_a^b g(x)dx$ diverges. Then

$$\int_a^b f(x)dx \geq \int_a^b g(x)dx = \infty$$

So $\int_a^b f(x)dx$ diverges

● **Comparison test, convergence version:** if you think that $\int_a^b f(x)dx$ converges, find a larger function whose integral also converges. That is, find a function g such that $f(x) \leq g(x)$ for all x in (a,b), and such that $\int_a^b g(x)dx$ converges. Then

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx < \infty$$

So $\int_a^b f(x)dx$ also converges.

As an alternative to the comparison test, there is the limit comparison test. This is useful when you can find a function which behaves just like the integrand near the problem spot

$f(x) \sim g(x)$ as $x \to a$ means the same thing as $\lim_{x\to a} \frac{f(x)}{g(x)} = 1$

The definition also applies if you replace both instances of $x \to a$ by $x \to \infty$ (or $x \to -\infty$). In any case, if your integrand f is really nasty and you can find a nicer function g such that $f(x) \sim g(x)$ as x approaches the problem spot, you're in business! That's because the limit comparison test says that whatever goes for g also goes for f. More precisely, here are two versions of the test depending on whether the problem spot is infinite or finite:

● **Limit comparison test, ∞ version:** find a simpler nonnegative function g with no problem spots in $[a,\infty)$, such that $f(x) \sim g(x)$ as $x \to \infty$. Then

- if $\int_a^\infty g(x)dx$ converges, so does $\int_a^\infty f(x)dx$; whereas
- if $\int_a^\infty g(x)dx$ diverges, so does $\int_a^\infty f(x)dx$

Of course, you can change the region $[a,\infty)$ into $(-\infty,b]$ and everything still works. There's also a version which applies when the problem spot is at some finite value a, which is at the left endpoint of the region of integration:

● **Limit comparison test, finite version:** find a simpler nonnegative function g with no problem spots in $(a, b]$ so that $f(x) \sim g(x)$ as $x \to a$. Then

- if $\int_a^b g(x)dx$ converges, so does $\int_a^b f(x)dx$; whereas

- if $\int_a^b g(x)dx$ diverges, so does $\int_a^b f(x)dx$.

Needless to say, this is also true if the only problem spot is at the right endpoint $x = b$ instead of $x = a$, provided that $f(x) \sim g(x)$ as $x \to b$ (not a).
So it's up to us to pluck an appropriate function g out of thin air to use as a comparison. It turns out that a lot of problems can be solved simply by taking $g(x)$ to be equal to $1/x^p$ for some appropriately chosen p. The convergence or divergence of the integral of such a function is precisely stated by the p-test:
● (**p-test**, $\int^\infty$ **version**) for any finite $a > 0$, the integral

$$\boxed{\int_a^\infty \frac{1}{x^p}dx \text{ converges if } p > 1 \text{ and diverges if } p \le 1}$$

● (**p-test**, $\int_0$ **version**) for any finite $a > 0$, the integral

$$\boxed{\int_0^a \frac{1}{x^p}dx \text{ converges if } p < 1 \text{ and diverges if } p \ge 1}$$

21.3 Behavior of Common Functions near ∞ and $-\infty$
OK, it's now time to answer the most important question of them all: how do you choose the comparison function g? This depends on whether the problem spot is at $\pm\infty, 0$, or some other finite value, so we'll consider these cases separately.
● Polynomials and poly-type functions near ∞ and $-\infty$
As far as polynomials are concerned, **the highest power dominates** as $x \to \infty$ or $x \to -\infty$. More precisely, suppose that p is a polynomial; then it's true that

If the highest-degree term of $p(x)$ is ax^n, then $p(x) \sim ax^n$ as $x \to \infty$ or as $x \to -\infty$

For example, we have (Page 456)

$$x^5 + 4x^4 + 1 \sim x^5 \quad \text{as } x \to \infty$$

If p is a poly-type function instead of a polynomial, a similar principle applies

$$3\sqrt{x} - 2\sqrt[3]{x} + 4 \sim 3\sqrt{x} \quad \text{as } x \to \infty$$

Since we have many new asymptotic relations, we can use the limit comparison test to analyze a lot of improper integrals. For example, consider

$$\int_0^\infty \frac{1}{x^5 + 4x^4 + 1}dx$$

We have

$$\frac{1}{x^5 + 4x^4 + 1} \sim \frac{1}{x^5} \quad \text{as } x \to \infty$$

Now, we have to be careful! We'd like to say that the integral we want behaves exactly like the integral $\int_0^\infty 1/x^5 dx$; the difficulty here is that this integral now has an extra problem spot at

$x=0$. In fact, this integral diverges, but only because of the problem spot at 0. This would lead to the wrong answer altogether. In order to avoid these inanities, we should have started by splitting the original integral into the pieces

$$\int_0^1 \frac{1}{x^5+4x^4+1}dx \quad \text{and} \quad \int_1^\infty \frac{1}{x^5+4x^4+1}dx$$

Beware of this situation-it arises often, so make sure that you split up the integral. Basically, if the "limit comparison function" g has a problem spot that the original function doesn't, you have to split up the original integral to avoid introducing a new problem spot. Normally the new integrand $g(x)$ will be of the form $1/x^p$, so you just need to avoid $x=0$ when you have a problem spot at ∞, just as in our example.

Finally, consider

$$\int_9^\infty \frac{1}{\sqrt{x^4+8x^3-9}-x^2}dx$$

As we discussed above, the highest power in the denominator is difficult to pin down, since $\sqrt{x^4}$ and $-x^2$ cancel out. So, we have to multiply top and bottom by the conjugate expression of the denominator (Page 459).

● Trig functions near ∞ and $-\infty$

Perhaps the only really useful thing we can say here is that

$$\boxed{|\sin(A)|\le 1} \quad \text{and} \quad \boxed{|\cos(A)|\le 1}$$

for **any** real number A. There are two main applications of the above inequalities. One is that you can use the comparison test in many cases. For example, does the integral

$$\int_5^\infty \frac{|\sin(x^4)|}{\sqrt{x}+x^2}dx$$

converge or diverge? Since the sine (or cosine) of **anything** is no more than 1 in absolute value. So, we have

$$\int_5^\infty \frac{|\sin(x^4)|}{\sqrt{x}+x^2}dx \le \int_5^\infty \frac{1}{\sqrt{x}+x^2}dx$$

And we can get that

$$\int_5^\infty \frac{|\sin(x^4)|}{\sqrt{x}+x^2}dx \le \int_5^\infty \frac{1}{\sqrt{x}+x^2}dx < \infty$$

The other nice application of the facts that $|\sin(A)|\le 1$ and $|\cos(A)|\le 1$ is that you can treat the sine or cosine of anything as inconsequential compared to any positive power of x, at least as $x\to\infty$ or $x\to-\infty$. For example

$$2x^3-3x^{0.1}+\sin(100x^{200}) \sim 2x^3 \quad \text{as } x\to\infty$$

● Exponentials near ∞ and $-\infty$

Here's a really useful principle: **exponentials grow faster than polynomials**. We first saw this in Section 9.4.4 of Chapter 9. There we expressed the principle in the form

$$\lim_{x\to\infty}\frac{x^n}{e^x}=0$$

where n is any positive number, even a very large one. Now consider the function f defined by $f(x)=x^n/e^x$ (Page 461). There must be some maximum height that the graph of $y=f(x)$ gets to. Let's call it C; this means that $f(x)=x^n/e^x\le C$ for all $x\ge 0$. (Note that you get a different C for each n, but that doesn't really affect us at all.) We can get the useful inequality

$$\boxed{e^{-x} \le \frac{C}{x^n} \quad \text{for all } x > 0}$$

The same is true if you replace e^{-x} by $e^{-p(x)}$, where $p(x)$ is any polynomial-type expression that goes to infinity when $x \to \infty$, and also if the base e is replaced by any other number greater than 1. The important point is that you get to choose any n you like, and you often have to be careful that you make it large enough. For example, consider

$$\int_1^\infty x^3 e^{-x} dx$$

We notice that

$$e^{-x} \le \frac{C}{x^5}$$

This is just the above boxed inequality, with n chosen to be 5. Why 5? Because it works:

$$\int_1^\infty x^3 e^{-x} dx \le \int_1^\infty x^3 \frac{C}{x^5} dx = C \int_1^\infty \frac{1}{x^2} dx < \infty$$

We have used the p-test to show that $C\int_1^\infty 1/x^2 dx$ converges. The comparison test now shows that the original integral converges as well. Now, how did I know to use x^5? What would happen if I used, say, $e^{-x} \le C/x^4$ instead? It doesn't work (Page 462). In practice, it's good to choose a number 2 more than the power you are trying to kill.

An important point: it is wrong, wrong, wrong to write $x^3 e^{-x} \sim e^{-x}$ as $x \to \infty$. It simply isn't true! If it were, then you could cancel out the positive quantity e^{-x} to conclude that $x^3 \sim 1$ as $x \to \infty$, and this is just crazy talk. So you should use the comparison test, not the limit comparison test, in the previous example.

How about e^x near $-\infty$? Well, this is the same thing as understanding the behavior of e^{-x} near ∞! For example, to investigate

$$\int_{-\infty}^{-4} x^{1000} e^x dx$$

Make $t = -x$. Since $dt = -dx$

$$\int_{-\infty}^{-4} x^{1000} e^x dx = -\int_\infty^4 (-t)^{1000} e^{-t} dt = \int_4^\infty t^{1000} e^{-t} dt$$

● Logarithms near ∞

First, notice that we don't consider logarithms near $-\infty$, because you can't take the log of a negative number! So it's futile to ask what happens to $\ln(x)$ as $x \to -\infty$.

On the other hand, logs grow slowly at ∞. In fact, they grow more slowly than any positive power of x. In symbols, we can say that if $\alpha > 0$ is some positive number of your choosing, then no matter how small it is, we have

$$\lim_{x\to\infty} \frac{\ln(x)}{x^\alpha} = 0$$

We looked at this principle in some detail in Section 9.4.5 of Chapter 9. By a similar argument to the one we used at the beginning of Section 21.3.3 above, you can show that there must a constant C such that

$$\boxed{\ln(x) \le Cx^\alpha \quad \text{for all } x > 1}$$

The same is true for logs of any base greater than 1, or if $\ln(x)$ is replaced by the log of a polynomial with positive leading coefficient.

For example, what do you make of (Page 465)

$$\int_2^\infty \frac{\ln(x)}{x^{1.001}}\,dx$$

The methodology is very similar to how we handled exponentials in Section 21.3.3 above.

Mind you, the principle that logs grow slowly isn't useful in every improper integral involving logs. Here are six improper integrals to consider: (Page 465)

$$\int_2^\infty \frac{\ln(x)}{x^{1.001}}\,dx, \quad \int_2^\infty \frac{1}{x^{1.001}\ln(x)}\,dx, \quad \int_2^\infty \frac{\ln(x)}{x}\,dx,$$

$$\int_2^\infty \frac{1}{x\ln(x)}\,dx, \quad \int_{3/2}^\infty \frac{\ln(x)}{x^{0.999}}\,dx, \quad \text{and} \quad \int_2^\infty \frac{1}{x^{0.999}\ln(x)}\,dx$$

21.4 Behavior of Common Functions near 0

We now know all about how polynomials, trig functions, exponentials, and logarithms behave at infinity. Now let's see what happens to them near zero.

● Polynomials and poly-type functions near 0

For polynomials, **the lowest power dominates** as $x \to 0$. This is the opposite of what happens as $x \to \infty$! To be more precise, suppose that p is a polynomial; then it's true that

$$\boxed{\text{if the lowest-degree term of } p(x) \text{ is } bx^m\text{, then } p(x) \sim bx^m \text{ as } x \to 0}$$

For example, $5x^4 - x^3 + 2x^2 \sim 2x^2$ as $x \to 0$.

For poly-type functions, it's not always easy to find the lowest-degree term, but the general principle still holds water. So, for example, $x^2 + \sqrt{x} \sim \sqrt{x}$ as $x \to 0^+$. The principle even works if constants are present-they are really multiples of x^0, which is a very low-degree term! So, for example, $2x^{1/3} + 4 \sim 4$ as $x \to 0$, as $4x^0$ has a lower exponent than $2x^{1/3}$.

● Trig functions near 0

Here are some very useful facts:

$$\boxed{\sin(x) \sim x, \quad \tan(x) \sim x, \quad \text{and} \quad \cos(x) \sim 1 \quad \text{as } x \to 0}$$

These are just restatements of limits we've already looked at in Chapter 7:

$$\lim_{x\to 0} \frac{\sin(x)}{x} = 1, \quad \lim_{x\to 0} \frac{\tan(x)}{x} = 1, \quad \text{and} \quad \lim_{x\to 0} \cos(x) = 1$$

Beware: these asymptotic relations only work with products and quotients, not sums and differences. For instance, you cannot write $\sin(x) - x \sim 0$ as $x \to 0$; see the end of Section 20.4.1.

Let's look at some examples (Page 470)

$$\int_0^1 \frac{1}{\tan(x)}\,dx \quad \text{and} \quad \int_0^1 \frac{1}{\sqrt{\tan(x)}}\,dx$$

A word of warning: just because we're looking at the behavior as $x \to 0$ doesn't mean that the problem spot has to be at 0. It might even be at ∞, as the following example shows:

$$\int_1^\infty \sin\left(\frac{1}{x}\right)dx$$

● Exponentials near 0

In some sense, **exponentials have no effect at $\mathbf{0}$**. More precisely,

$$\boxed{e^x \sim 1 \quad \text{and} \quad e^{-x} \sim 1 \quad \text{as } x \to 0}$$

This is just another way of saying that

$$\lim_{x\to 0} e^x = 1 \quad \text{and} \quad \lim_{x\to 0} e^{-x} = 1$$

For example, the improper integral

$$\int_0^1 \frac{e^x}{x\cos(x)}\,dx$$

diverges, because

$$\frac{e^x}{x\cos(x)} \sim \frac{1}{x\cdot 1} = \frac{1}{x} \quad \text{as } x \to 0$$

Beware: this only applies to the exponential of a small quantity (like x or $-x$). An example of a tricky integral where you could trip up is

$$\int_0^1 \frac{e^{-1/x}}{x^5}\,dx$$

It would be wrong to write $e^{-1/x} \sim 1$, since $1/x \to \infty$ as $x \to 0^+$.

Here's another possible trap. In the integral

$$\int_0^2 \frac{dx}{\sqrt{e^x - 1}}$$

We need to be cleverer, from Section 9.4.2 of Chapter 9 to conclude that

$$\boxed{e^x - 1 \sim x \quad \text{as } x \to 0}$$

It follows that

$$\frac{1}{\sqrt{e^x - 1}} \sim \frac{1}{\sqrt{x}} \quad \text{as } x \to 0^+$$

● Logarithms near 0

Here the principle is that logs go to $-\infty$ slowly as $x \to 0^+$. Let's make things go to ∞ instead by taking absolute values, remembering that $\ln(x)$ is negative when $0 < x < 1$. So the idea is that no matter how small $\alpha > 0$ is, there's some constant C such that

$$\boxed{|\ln(x)| \le \frac{C}{x^\alpha} \quad \text{for all } 0 < x < 1}$$

This follows from the limit

$$\lim_{x\to 0^+} x^\alpha \ln(x) = 0$$

which we looked at in Section 9.4.6 of Chapter 9.

So, to understand (Page 473)

$$\int_0^1 \frac{|\ln(x)|}{x^{0.9}}\,dx$$

● The behavior of more general functions near 0

In Section 24.2.2 of Chapter 24, we'll learn about **Maclaurin series**. Anyway, the basic idea is that if a function has a Maclaurin series which converges to the function near 0, then the function is asymptotic to the lowest-order term in the series as $x \to 0$. That is

$$\boxed{\text{if } f(x) = a_n x^n + a_{n+1}x^{n+1} + \cdots, \text{then } f(x) \sim a_n x^n \text{ as } x \to 0}$$

Consider the following examples:

$$\int_0^1 \frac{dx}{1-\cos(x)} \quad \text{and} \quad \int_0^1 \frac{dx}{(1-\cos(x))^{1/3}}$$

21.5 How to Deal with Problem Spots Not at 0 or ∞

If a problem spot occurs at some finite value other than 0, do a substitution. Specifically:

● If the only problem spot in $\int_a^b f(x)dx$ occurs at $x=a$, make the substitution $t=x-a$.

Note that $dt=dx$. The new integral has a problem spot at 0 only

● If the only problem spot in $\int_a^b f(x)dx$ occurs at $x=b$, make the substitution $t=b-x$.

Note that $dt=-dx$. Use the minus sign to switch the limits of integration. The new integral should have a problem spot at 0 only.

CHAPTER 22 Sequences and Series: Basic Concepts

Here's the good news: infinite series are pretty similar to improper integrals. So a lot, but not all, of the relevant techniques are shared and we don't need to reinvent the wheel. In order to define what an infinite series is, we'll also need to look at sequences.

22.1 Convergence and Divergence of Sequences

A **sequence** is a collection of numbers in order. It might have a finite number of terms, or it might go on forever, in which case it is called an *infinite sequence*. For example,

$$0, 1, -1, 2, -2, 3, -3, \cdots$$

is an infinite sequence which incidentally includes every integer, positive and negative. Sequences are normally written using subscript notation, where a_1 denotes the first element of the series, a_2 the second, a_3 the third, and so on. (Sometimes a_0 is the first element, a_1 the second, and so on. Also, we don't have to use a; for example, b_n or any other letter is fair game.)

Given an infinite sequence, our main focus is going to be on the limiting behavior of the values of the sequence as the index n tends to infinity. That is, what happens to the sequence as you look farther and farther along it? In math notation, does

$$\lim_{n\to\infty} a_n$$

exist, and if so, what is it? By the way, we haven't really defined the above limit, but the definition is not much different from the definition of $\lim\limits_{x\to\infty} f(x)$ for a function f. The basic idea is that the statement

$$\lim_{n\to\infty} a_n = L$$

means that a_n might wander around for a little while, but eventually gets very close-as close as you like-to L and stays at least as close to L for ever after. If there's such a number L, then the sequence $\{a_n\}$ *converges*; otherwise it *diverges*.

By the way, as we did with functions, we sometimes say that $a_n \to L$ as $n \to \infty$. This means the same thing as saying $\lim\limits_{n\to\infty} a_n = L$.

● The connection between sequences and functions

Consider the sequence given by

$$a_n = \frac{\sin(n)}{n^2}$$

which we looked at earlier. This is closely related to the function f defined by

$$f(x) = \frac{\sin(x)}{x^2}$$

In fact, a_n is equal to $f(n)$ for each positive integer n. So if we can establish that $\lim\limits_{x\to\infty} f(x)$ exists, then we'll know that the sequence $\{a_n\}$ has the same limit. The sequence inherits the limiting properties of the function. There's also a connection to horizontal asymptotes $y = L$.

Inspired by these observations, we can easily extend some other properties of limits of functions to the case of sequences. For example, if you have two convergent sequences $\{a_n\}$ and $\{b_n\}$, such that $a_n \to L$ and $b_n \to M$ as $n \to \infty$, then the sum $a_n + b_n$ gives a new sequence which converges to $L+M$. The same goes for differences, products, quotients (provided that $M \neq 0$, since you can't divide by 0), and constant multiples.

Another useful fact is that the sandwich principle: In math-speak, if $c_n \leq a_n \leq b_n$ and both $b_n \to L$ and $c_n \to L$ as $n \to \infty$, then $a_n \to L$ as $n \to \infty$ as well (Page 479).

Another property which transfers over from functions is that **continuous functions respect limits**. Well, suppose that $a_n \to L$ as $n \to \infty$. Then if f is a function which is continuous at $x = L$, we can say that $f(a_n) \to f(L)$ as $n \to \infty$. The limit relation is preserved when you hit everything with f (Page 479).

One more useful tool that we can borrow from the theory of functions is l'Hôpital's Rule. The problem with using the rule on a sequence is that you can't differentiate the quantity a_n with respect to the variable n, since n has to be an integer. Indeed, when you differentiate a function f with respect to a variable x, the idea is that you wobble x around a little and see what happens to $f(x)$. You can't wobble an integer around because it wouldn't be an integer any more. So, if you want to use l'Hôpital's Rule, you have to embed the sequence in a suitable function first. For example, if $a_n = \ln(n)/\sqrt{n}$, you can find $\lim\limits_{n\to\infty} a_n$ by letting (Page 480)

$$f(x) = \frac{\ln(x)}{\sqrt{x}}$$

● Two important sequences

Pick some constant number r and consider the sequence given by $a_n = r^n$ starting at $n = 0$. This is a ***geometric progression***. Notice that each term is a constant multiple of the previous one. Let's look at a few examples of geometric progressions:

- if $r = 0$, the sequence is just $0, 0, 0, \cdots$, which clearly converges to 0
- if $r = 1$, the sequence is just $1, 1, 1, \cdots$, which clearly converges to 1
- if $r = 2$, the sequence is just $1, 2, 4, 8, \cdots$, which evidently converges to ∞
- if $r = -1$, the sequence is just $1, -1, 1, -1, 1, \cdots$, which diverges, but not to ∞ or $-\infty$, because it keeps on oscillating back and forth between -1 and 1- in other words, the limit does not exist (DNE)
- if $r = -2$, the sequence is just $1, -2, 4, -8, \cdots$, which diverges in the same way (the limit does not exist)-in fact, this time the oscillations are even wilder
- if $r = 1/2$, the sequence is just $1, 1/2, 1/4, 1/8, \cdots$, which converges to 0; and finally
- if $r = -1/2$, the sequence is just $1, -1/2, 1/4, -1/8, \cdots$, which also converges to 0, despite the oscillations, since these oscillations eventually become as small as you like

These are all special cases of the general rule, which is as follows:

$$\boxed{\lim_{n\to\infty} r^n \begin{cases} = 0 & \text{if } -1 < r < 1 \\ = 1 & \text{if } r = 1 \\ = \infty & \text{if } r > 1 \\ \text{DNE} & \text{if } r \leq -1 \end{cases}}$$

Geometric progressions don't have to start at 1. If we set $a_n = ar^n$, where a is some constant, then the first term a_0 is equal to a. Most important, if $-1 < r < 1$, then $\lim\limits_{n\to\infty} ar^n$ is 0 regardless of the value of a.

Let's look at the limit of another sequence very quickly. In particular, if k is any constant, then

$$\boxed{\lim_{n\to\infty}\left(1+\frac{k}{n}\right)^n = e^k}$$

22.2 Convergence and Divergence of Series

A **series** is just a sum. We'd like to add up all of the terms of a sequence a_n.

You begin with an infinite **sequence**

$$\{a_n\} = a_1, a_2, a_3, \cdots$$

and use it to construct an infinite **series**:

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \cdots$$

To understand the limiting behavior of this series, make a **new sequence** of partial sums:

$$A_N = \sum_{n=1}^{N} a_n = a_1 + a_2 + a_3 + \cdots + a_{N-1} + a_N$$

By definition, the limit of the series is the same as the limit of the new sequence of partial sums, if the limit exists; otherwise the series diverges.

Here's how to define the value of an infinite series using sigma notation: (Page 483)

$$\boxed{\sum_{n=1}^{\infty} a_n = \lim_{N\to\infty}\sum_{n=1}^{N} a_n}$$

By the way, we don't need to begin our series at $n = 1$. You can begin at any number, even $n = 0$. Now, here's an important point: whether a series converges or diverges has nothing to do with the starting point of the series!

● Geometric series (theory)

Let's look at an important example of an infinite series. Suppose we start with the geometric progression $1, r, r^2, r^3, \cdots$, which we looked at in Section 22.1.2 above. We can use this sequence as the terms of an infinite series:

$$1 + r + r^2 + r^3 + \cdots = \sum_{n=0}^{\infty} r^n$$

This is called a ***geometric series***. The question is, does it converge, and if so, to what?

To find out, we'd better look at the partial sums (Page 484). Hopefully, in your previous math studies you've seen that the above expression can be simplified as follows:

$$A_N = 1 + r + r^2 + r^3 + \cdots + r^{N-1} + r^N = \frac{1 - r^{N+1}}{1 - r}$$

First, suppose that $-1 < r < 1$. Then we saw in (Page 485). Here's how the whole argument looks on one line, using sigma notation:

$$\sum_{n=0}^{\infty} r^n = \lim_{N\to\infty}\sum_{n=0}^{N} r^n = \lim_{N\to\infty}\frac{1 - r^{N+1}}{1 - r} = \frac{1}{1 - r}$$

How about when r isn't between -1 and 1? It turns out that the geometric series must diverge in this case; we'll see why at the end of the next section. So, in summary:

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r} \quad \text{if} -1 < r < 1$$

otherwise, if $r \geq 1$ or $r \leq -1$, the series diverges

In the above geometric series, the first term is always 1, since $r^0 = 1$. If you start at some other number a instead, then the terms are a, ar, ar^2, and so on. So you can get a more general form of the above principle:

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r} \quad \text{if} -1 < r < 1$$

otherwise, if $r \geq 1$ or $r \leq -1$, the series diverges

22.3 The nth Term Test (Theory)

For a series to converge, the sequence of partial sums has to have a limit. So, your step sizes, which are just given by the sequence $\{a_n\}$, eventually have to become very small, at least if you want your series to converge. Mathematically, this means that you need to have $a_n \to 0$ as $n \to \infty$. This leads us to the *nth term test*:

nth term test: if $\lim_{n\to\infty} a_n \neq 0$, or the limit doesn't exist, then the series $\sum_{n=1}^{\infty} a_n$ diverges

If $\lim_{n\to\infty} a_n = 0$, then the series may converge or it may diverge, and you have to do more work to resolve the issue. Just beware: **the nth term test cannot be used to show that a series converges**!

In a convergent series, although the terms a_n must go to 0, that doesn't mean that the limit of the series is 0.

22.4 Properties of Both Infinite Series and Improper Integrals

It turns out that there are some connections between infinite series and improper integrals, particularly improper integrals with a problem spot at ∞.

● The comparison test (theory)

This is basically the same as the comparison test for improper integrals (Page 487).

For example, consider

$$\sum_{n=3}^{\infty} \left(\frac{1}{2}\right)^3 |\sin(n)|$$

Using the comparison test

$$\sum_{n=3}^{\infty} \left(\frac{1}{2}\right)^3 |\sin(n)| \leq \sum_{n=3}^{\infty} \left(\frac{1}{2}\right)^3 < \infty$$

● The limit comparison test (theory)

In Section 20.4.1 of Chapter 20, we made the following definition:

$f(x) \sim g(x)$ as $x \to \infty$ means the same thing as $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$

There's a version of this for sequences that looks almost the same:

$$\boxed{a_n \sim b_n \text{ as } n \to \infty \text{ means the same thing as } \lim_{n\to\infty} \frac{a_n}{b_n} = 1}$$

By the way, if $a_n \sim b_n$ as $n \to \infty$, we say that the sequences are *asymptotic* to each other.
For example, consider

$$\sum_{n=0}^{\infty} \sin\left(\frac{1}{2^n}\right)$$

We see that

$$\sin\left(\frac{1}{2^n}\right) \sim \frac{1}{2^n} \quad \text{as } \frac{1}{2^n} \to 0$$

So the above relation can be written as

$$\sin\left(\frac{1}{2^n}\right) \sim \left(\frac{1}{2}\right)^n \quad \text{as } n \to \infty$$

● The p-test (theory)

There's also a p-test for series. It's basically the same as the p-test for improper integrals with problem spot at 1. In particular, it says that

$$\sum_{n=a}^{\infty} \frac{1}{n^p} \quad \begin{cases} \text{converges} & \text{if } p > 1 \\ \text{diverges} & \text{if } p \le 1 \end{cases}$$

Some simple examples of the p-test are that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \text{ converges, but } \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \text{ diverges}$$

● The absolute convergence test

If the series keeps switching between positive and negative terms? Some examples of this are

$$\sum_{n=3}^{\infty} \sin(n)\left(\frac{1}{2}\right)^n, \quad \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}, \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{(-1)^n}{n}$$

The second and third of these series are actually *alternating series*. This means that the terms alternate between positive and negative numbers.

The absolute convergence test says that if $\sum_{n=1}^{\infty} |a_n|$ converges, so does $\sum_{n=1}^{\infty} a_n$. Again, the series can start at any value of n, not necessarily $n = 1$.

22.5 New Tests for Series

Let's look at four tests for convergence of series which have no corresponding improper integral version: the ratio test, the root test, the integral test, and the alternating series test.

● The ratio test (theory)

Here's a really really useful test which only works for series, not improper integrals. It's called the *ratio test* because it involves the ratio of successive terms of a sequence.

Let's set the scene: suppose we have a series $\sum_{n=1}^{\infty} a_n$. We'd like the terms to go to 0 fast enough for this series to converge. Here's one way this can happen: suppose we consider a new sequence, which we'll call b_n, of the absolute value of ratios of successive terms of the series. That is, we let

$$b_n = \left|\frac{a_{n+1}}{a_n}\right|$$

for each n. This is a sequence, so maybe it converges to something. Now here's the result: if the sequence $\{b_n\}$ converges to a number less than 1, then we can immediately conclude that the series $\sum_{n=1}^{\infty} a_n$ converges. In fact, it converges absolutely: that is, $\sum_{n=1}^{\infty} |a_n|$ also converges. On the other hand, if the sequence $\{b_n\}$ converges to a number greater than 1, then the series $\sum_{n=1}^{\infty} a_n$ diverges. If the sequence $\{b_n\}$ converges to 1, or if it doesn't converge, then we can't say anything about the original series.

● The root test (theory)

The root test (also called the nth root test) is a close cousin of the ratio test. Instead of considering ratios of successive terms, just consider the nth root of the absolute value of the nth term. That is, starting with a series $\sum_{n=1}^{\infty} a_n$, let's make a new sequence given by

$$b_n = |a_n|^{1/n}$$

Now you see whether the sequence $\{b_n\}$ converges and try to find the limit. If the limit is less than 1, then the series $\sum_{n=1}^{\infty} a_n$ converges (in fact, converges absolutely). If the limit is greater than 1, the series diverges. If the limit equals 1, then you can't tell what the heck is going on and have to try something else.

● The integral test (theory)

We already saw in Section 22.4 above that there's a connection between improper integrals and infinite series. The integral test really nails down this connection.

In summary, we have the **integral test**: if f is a decreasing positive function such that $f(n) = a_n$ for all positive integers n, then

$$\int_1^{\infty} f(x)dx \quad \text{and} \quad \sum_{n=1}^{\infty} a_n$$

either both converge or both diverge. Again, the series can start at any number, not just $n = 1$; Just change the lower bound of the integral to match.

● The alternating series test (theory)

When a series converges, but its absolute version diverges, we say that the series *converges conditionally*. So $\sum_{n=1}^{\infty}(-1)^n/n$ converges conditionally. Let's see why.

The *alternating series test* says that if a series $\sum_{n=1}^{\infty} a_n$ is alternating, and the absolute values of its terms are decreasing to 0, then the series converges. That is, we need a_n to be alternately positive and negative, and $|a_n|$ to be decreasing, and $\lim_{n\to\infty}|a_n| = 0$. In that case, the series converges.

CHAPTER 23 How to Solve Series Problems

The scenario: you are given a series $\sum_{n=1}^{\infty} a_n$, and you want to know whether or not it converges. If it does converge, then perhaps you'd like to know its value (that is, what it converges to). The series has to be pretty special in order to find a nice expression for its value. Of course, the series may not start at $n = 1$ as in the above series-it could be $n = 0$ or some other value of n.

This chapter is all about giving you a blueprint of how to proceed. Here's a possible flowchart for how to approach a series:

1. **Is the series geometric?** If your series only involves exponentials like 2^n or e^{3n}, it might be a geometric series, or it might be the sum of one or more geometric series
2. **Do the terms go to 0?** If the series isn't geometric, try the **nth term test**. Check that the terms converge to 0; otherwise the series diverges by the nth term test
3. **Are there negative terms in the series?** If so, you may have to use the **absolute convergence test** or the **alternating series test**
4. **Are factorials involved?** If so, use the **ratio test**. The test is also useful when there are exponentials involved but the series isn't geometric
5. **Are there tricky exponentials with n in the base and the exponent?** If so, try the **root test**. In general, if it is easy to take the nth root of the term a_n, the root test is probably a winner
6. **Do the terms have a factor of exactly $1/n$ as well as logarithms?** In that case, the **integral test** is probably what you want
7. **Do none of the above tests seem to work?** You may have to use the **comparison test** or the **limit comparison test** in conjunction with the **p-test**.

The above blueprint will help guide your way through a lot of different series. It's not perfect! There are always tricks and traps that could arise.

23.1 How to Evaluate Geometric Series

If your series only involves exponentials like 2^n or e^{3n}, it might be the sum of one or more geometric series. As we saw in the previous chapter, geometric series are simple enough that you can actually find their values (if they converge). The general form of a geometric series is $\sum_{n=m}^{\infty} ar^n$, where r is the common ratio. Rather than learn the formula in mathematical language, I recommend learning it in words:

$$\boxed{\text{sum of infinite geometric series} = \frac{\text{first term}}{1-\text{common ratio}}, \text{ if } -1 < \text{ratio} < 1}$$

If the common ratio isn't between -1 and 1, then the series diverges (Page 502).

23.2 How to Use the nth Term Test

Always try the nth term test first! The test says:

$$\boxed{\text{if } \lim_{n\to\infty} a_n \neq 0, \text{ or the limit doesn't exist, then the series } \sum_{n=1}^{\infty} a_n \text{ diverges}}$$

If the terms of your series don't tend to 0, the series must diverge. If the terms do tend to 0, the series might converge or it might diverge: you have to do more work. **This test cannot be used to show that a series converges**.

23.3 How to Use the Ratio Test

Use the ratio test whenever factorials are involved. Remember, factorials involve exclamation points, such as in $n!$ or $(2n+5)!$. The ratio test is also often useful when there are exponentials around, such as 2^n or $(-5)^{3n}$. Here's the statement of the test

> if $L = \lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right|$, then $\sum_{n=1}^{\infty} a_n$ converges absolutely if $L < 1$,
>
> and diverges if $L > 1$; but if $L = 1$ or the limit doesn't exist, then the ratio test tells you nothing

To use the ratio test, always start with the following framework:

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty}\left|\frac{n\text{th term with } n \text{ replaced by } (n+1)}{n\text{th term}}\right|$$

1. If $L < 1$, then the original series $\sum_{n=1}^{\infty} a_n$ converges; in fact, it converges absolutely
2. If $L > 1$, then the original series diverges
3. If $L = 1$, or the limit doesn't exist, then the ratio test is useless. Try something else

23.4 How to Use the Root Test

Use the root test when there are a lot of tricky exponentials around involving functions of $\boldsymbol{n}$. It's especially useful when the terms of your series look like A^B, where both A and B are functions of n. Here's the statement of the test

> if $L = \lim_{n\to\infty}|a_n|^{1/n}$, then $\sum_{n=1}^{\infty} a_n$ converges absolutely if $L < 1$,
>
> and diverges if $L > 1$; but if $L = 1$ or the limit doesn't exist, then the ratio test tells you nothing

To use the root test, always start off with the following expression:

$$\lim_{n\to\infty}|a_n|^{1/n}$$

and then replace a_n by the general term of the series. Find the limit (if it exists) and call it L. Then you have three possibilities, which are identical to the possibilities which arise in the ratio test. The conclusions are luckily the same as well:

1. If $L < 1$, then the original series $\sum_{n=1}^{\infty} a_n$ converges; in fact, it converges absolutely
2. If $L > 1$, then the original series diverges
3. If $L = 1$, or the limit doesn't exist, then the root test is useless. Try something else

23.5 How to Use the Integral Test

Use the integral test when the series involves both $\boldsymbol{1/n}$ and $\boldsymbol{\ln(n)}$. If N is any positive integer, then we can say:

> if $a_n = f(n)$, for some continuous decreasing function f, then $\sum_{n=N}^{\infty} a_n$ and $\int_N^{\infty} f(x)dx$ either both converge or both diverge

In practice, here are the steps involved in using the integral test

- Replace n by x, change $\sum_{n=1}^{\infty}$ into $\int_1^{\infty}$, and put a dx at the end. Of course, if the series begins at $n = 2$, then you use $\int_2^{\infty}$ instead, for example
- Check that the integrand is decreasing; you can do that by showing that the derivative is negative, or just by inspecting the integrand directly

● Now deal with the improper integral from the first step. The main advantage of integrals over series is that you can use a substitution (or change of variables, if you prefer) in an integral. The most common substitution in this context is $t = \ln(x)$

● If the improper integral converges, so does the series. If the integral diverges, the series diverges too.

23.6 How to Use the Comparison Test, the Limit Comparison Test, and the p-test

Use these tests for series with positive terms when none of the other tests seem to apply. You definitely want to try the nth term test first, then use the ratio test if factorials are involved, the root test if the terms have exponentials where the base and exponent are both functions of n, or the integral test if you have a factor of $1/n$ and logarithms are involved. What does that leave? Basically the same tools as you have for integrals: the comparison test, the limit comparison test, the p-test, and an understanding of how common functions behave near ∞ and near 0. In any case, here are the tests once more. (For the comparison and limit comparison tests, we assume all the terms a_n are nonnegative.)

1. **Comparison test, divergence version:** if you think $\sum_{n=1}^{\infty} a_n$ diverges, find a smaller series which also diverges. That is, find a positive sequence $\{b_n\}$ such that $a_n > b_n$ for all n, and such that $\sum_{n=1}^{\infty} b_n$ diverges. Then

$$\sum_{n=1}^{\infty} a_n > \sum_{n=1}^{\infty} b_n = \infty$$

so $\sum_{n=1}^{\infty} a_n$ diverges

2. **Comparison test, convergence version:** if you think $\sum_{n=1}^{\infty} a_n$ converges, find a larger series which also converges. That is, find $\{b_n\}$ such that $a_n \leq b_n$ for all n, and such that $\sum_{n=1}^{\infty} b_n$ converges. Then

$$\sum_{n=1}^{\infty} a_n \leq \sum_{n=1}^{\infty} b_n < \infty$$

so $\sum_{n=1}^{\infty} a_n$ converges

3. **Limit comparison test:** find a simpler series $\sum_{n=1}^{\infty} b_n$ so that $a_n \sim b_n$ as $n \to \infty$. Then if $\sum_{n=1}^{\infty} b_n$ converges, so does $\sum_{n=1}^{\infty} a_n$. On the other hand, if $\sum_{n=1}^{\infty} b_n$ diverges, then so does $\sum_{n=1}^{\infty} a_n$. (Remember that "$a_n \sim b_n$ as $n \to \infty$" means the same thing as "$\lim\limits_{n\to\infty} a_n / b_n = 1$")

4. **p-test:** if $a \geq 1$, the series

$$\sum_{n=a}^{\infty} \frac{1}{n^p} \begin{cases} \text{converges} & \text{if } p > 1 \\ \text{diverges} & \text{if } p \leq 1 \end{cases}$$

This is the same as the $\int^{\infty}$ version of the p-test for integrals.

Now consider the series (Page 513)

$$\sum_{n=1}^{\infty} \sin\left(\frac{1}{n}\right)$$

A really nasty series: (Page 513)

$$\sum_{n=2}^{\infty} \cos^2(n) \tan\left(\frac{(n^2 + 4n - 3)\ln(n)}{\sqrt{n^7 + 2n^4 + 3n}}\right)$$

23.7 How to Deal with Series with Negative Terms

Suppose some of the numbers a_n which appear as terms in your series are negative. Here are some ways to handle this situation:

1. **If all the terms a_n are negative, then modify the series by putting a minus sign in front of all the terms**
2. **If some terms are positive and some terms are negative, try the nth term test first**
3. **If some terms are positive and some terms are negative, and the terms converge to 0 as $n \to \infty$, next try the absolute convergence test**:

if $\sum_{n=1}^{\infty}\lvert a_n\rvert$ converges, then so does $\sum_{n=1}^{\infty} a_n$

In this case, we say that the sequence is *absolutely convergent* or that it converges absolutely

4. **If the series doesn't converge absolutely, try the alternating series test**. As we saw in Section 22.5.4 of the previous chapter

if the absolute values of the terms of an alternating series decrease to 0 monotonically as $n \to \infty$, the series converges

So there are actually three things to check if you want to use the test on a series $\sum_{n=1}^{\infty} a_n$:

● the terms a_n alternate between positive and negative (that is, the signs of the terms are, in order, $+,-,+,-,\cdots$, or perhaps $-,+,-,+,\cdots$)

● the quantities $|a_n|$ tend to 0 as n gets large; that is,

$$\lim_{n\to\infty}|a_n| = 0$$

● the absolute values of the terms $|a_n|$ are decreasing in n (so the underlying sequence is getting smaller and smaller, in terms of absolute value).

If all three of these properties are true, then the series converges. **Note: you should always try the absolute convergence test first. If the series converges absolutely, do not use the alternating series test!**

CHAPTER 24 Introduction to Taylor Polynomials, Taylor Series, and Power Series

We now come to the important topics of power series and Taylor polynomials and series.

24.1 Approximations and Taylor Polynomials

Here's a nice fact: for any real number x, we have

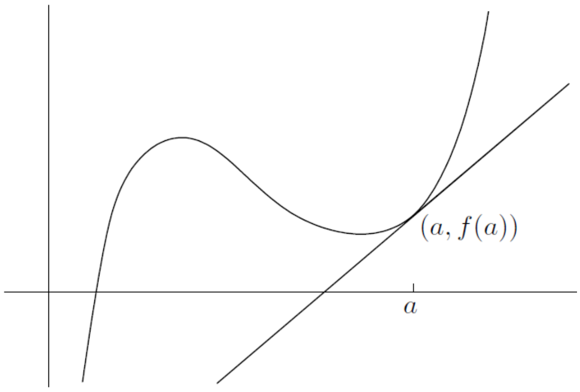
$$e^x \cong 1 + x + \frac{x^2}{2} + \frac{x^3}{3}$$

Also, the closer x is to 0, the better the approximation.

● Linearization revisited

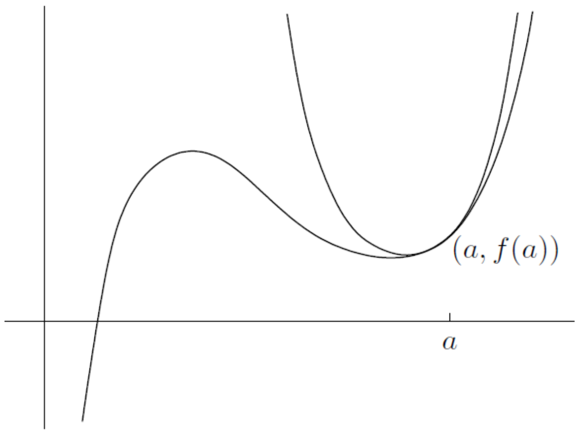
What is the equation of the line which best approximates the curve $y = f(x)$ near the point $(a, f(a))$? The answer to this question is that the line we're looking for is the tangent line to the curve at the point $(a, f(a))$, and its equation is

$$y = f(a) + f'(a)(x - a)$$



● Quadratic approximations

Why stick to lines, though? Let's ask the same question we did at the beginning of the previous section, but with parabolas instead. Here is our question: what is the equation of the quadratic which best approximates the curve $y = f(x)$ near $(a, f(a))$? Using the same function as in the picture above, here's a guess as to what the quadratic should look like:



It turns out that the formula for the quadratic which best approximates the curve $y = f(x)$ for x near a (that is, near the point $(a, f(a))$ on the curve) is given by

$$y = f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2$$

You can think of the last term $\frac{1}{2}f''(a)(x-a)^2$ as a so-called *second-order correction term*. This means that we should actually be able to do a better job of approximation than just by using the tangent line. The second-order correction term helps us get even closer to the curve, at least for x near a.

● Higher-degree approximations

A Taylor approximation theorem: if f is smooth at $x=a$, then of all the polynomials of degree N or less, the one which best approximates $f(x)$ for x near a is given by

$$P_N(x)=f(a)+f'(a)(x-a)+\frac{f''(a)}{2!}(x-a)^2$$
$$+\frac{f^{(3)}(a)}{3!}(x-a)^3+\cdots+\frac{f^{(N)}(a)}{N!}(x-a)^N$$

In sigma notation, the formula looks like this:

$$P_N(x)=\sum_{n=0}^{N}\frac{f^{(n)}(a)}{n!}(x-a)^n$$

In this formula, remember that $0!=1$, that $f^{(0)}(a)$ means the same thing as $f(a)$ (zero derivatives), and that $f^{(1)}(a)$ means the same thing as $f'(a)$ (one derivative).

We call the polynomial P_N the *Nth-order Taylor polynomial of* $f(x)$ at $x=a$ (Page 522).

Once again, the important property of P_N is that

$$P_N^{(n)}(a)=f^{(n)}(a)$$

for all $n=0,1,\cdots,N$.

The Taylor approximation theorem actually depends on Taylor's Theorem, which we'll look at in the next section.

● Taylor's Theorem

Here's what the error is in our case:

$$\text{error}=\text{true value}-\text{approximate value}=e^{-\frac{1}{10}}-\frac{5429}{6000}$$

What we'd really like is another formula for the error. That's where Taylor's theorem comes in. We want to use the value of $P_N(x)$ to approximate the true value $f(x)$, so we consider the error term, which is the difference between the true value and the approximate value:

$$R_N(x)=f(x)-P_N(x)$$

Actually, $R_N(x)$ is called the *Nth-order error term*; it's also referred to as the *Nth-order remainder term*, since it's all that remains when you take $P_N(x)$ away from $f(x)$. As promised above, Taylor's Theorem gives an alternative formula for $R_N(x)$:

Taylor's Theorem: the Nth-order remainder term by $R_N(x)$ about $x=a$ is

$$R_N(x)=\frac{f^{(N+1)}(c)}{(N+1)!}(x-a)^{N+1}$$

where c is some number which lies between x and a.

Note that the number c depends on what x and N are, and cannot be determined in general!

Since $f(x)=P_N(x)+R_N(x)$, we can write the whole kit and caboodle as

$$\boxed{f(x)=\sum_{n=0}^{N}\frac{f^{(n)}(a)}{n!}(x-a)^n+\frac{f^{(N+1)}(c)}{(N+1)!}(x-a)^{N+1}}$$

This seems pretty nasty. And what on earth is with this number c, anyway? (Page 524)

24.2 Power Series and Taylor Series

Here's another fact:

$$e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\frac{x^5}{5!}+\cdots$$

for all real numbers x. You might notice that it looks similar to the approximation at the beginning of Section 24.1 above, but there are two big differences. First, we're no longer dealing with an approximation, and second, there's an infinite series on the right-hand side!

So, let's see if we can understand what the above equation actually means. Suppose we start with the right-hand side,

$$1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\frac{x^5}{5!}+\cdots$$

This looks like a polynomial, but it isn't, since there's no highest-degree term. It just keeps on going forever. In fact, it's an example of a power series.

If $x=-1/10$, you get the series

$$1-\frac{1}{10}+\frac{1/100}{2!}-\frac{1/1000}{3!}+\frac{1/10000}{4!}-\frac{1/100000}{5!}+\cdots$$

This series might converge, or it might diverge. So which is it? The answer is that it converges, and what's more, we even know that it converges to $e^{-1/10}$.

● Power series in general

A *power series about* $x=0$ is an expression of the form

$$a_0+a_1x+a_2x^2+a_3x^3+a_4x^4+\cdots$$

where the numbers a_n are fixed constants. Even though a power series isn't a polynomial, we'll still refer to a_n as the *coefficient* of x^n in the power series. The above series can also be written using sigma notation as

$$\sum_{n=0}^{\infty}a_nx^n$$

Something nice happens to the power series

$$a_0+a_1x+a_2x^2+a_3x^3+a_4x^4+\cdots$$

when you set $x=0$: all the terms vanish except for the a_0 at the beginning, so the series automatically converges (to a_0, of course!).

Let's transfer this special property over to some other number a. All we have to do is replace x by $(x-a)$. So here is the general expression for a *power series* about $x=a$:

$$\boxed{a_0+a_1(x-a)+a_2(x-a)^2+a_3(x-a)^3+a_4(x-a)^4+\cdots}$$

In sigma notation, this looks like

$$\boxed{\sum_{n=0}^{\infty}a_n(x-a)^n}$$

This series converges for sure when $x = a$, since all the terms except a_0 vanish. The number a is called the *center of the power series*.

● Taylor series and Maclaurin series

In the previous section, we saw that a general power series about $x = a$ is given (using sigma notation and also in expanded form) by

$$\sum_{n=0}^{\infty} a_n(x-a)^n = a_0 + a_1(x-a) + a_2(x-a)^2 + a_3(x-a)^3 + a_4(x-a)^4 + \cdots$$

This converges for $x = a$, and might converge for other values of x. We could then plug in all these values of x one at a time, find what the series converges to in each case, and call that $f(x)$. So, starting with a power series, we have defined a function.

Suppose that we instead start off with some smooth function f. We're going to define a special power series about $x = a$ by using all the derivatives of f:

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

When you expand the sigma notation, this becomes

$$f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \frac{f^{(4)}(a)}{4!}(x-a)^4 + \cdots$$

The coefficients of this power series are given by $a_n = f^{(n)}(a)/n!$. The series is called the ***Taylor series*** of f about $x = a$. So, starting with a function, we have defined a power series.

In other words, the Taylor polynomial $P_N(x)$ is the Nth partial sum of the Taylor series.

We have just one more definition: the ***Maclaurin series*** *of* f is just another name for the Taylor series of f about $x = 0$. So it's given by

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n$$

● Convergence of Taylor series

OK, let's review the situation. We started out with a function f and a number a, and we constructed the Taylor series of f about $x = a$

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

since then we'd know that the Taylor series converges for any x and also that it converges to the original function value $f(x)$. The problem is, the above equation isn't always valid (Page 530).

So, how do you know if and when a Taylor series actually converges to its underlying function? Start by writing (Page 531)

$$f(x) = P_N(x) + R_N(x)$$

In other words, **if you want to prove that a function equals its Taylor series at some number x, try to show that $R_N(x) \to 0$ as $N \to \infty$**

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$

24.3 A Useful Limit

This section isn't about power series at all-it just contains a proof of a limit we needed twice in the previous section: (Page 534)

$$\lim_{N\to\infty} \frac{x^{N+1}}{(N+1)!} = 0$$

Hint: Think of it like a substitution, $n = N + 1$, it likes

$$\lim_{n\to\infty} \frac{x^n}{n!} = 0$$

To prove that the series

$$\sum_{n}^{\infty} \frac{x^n}{n!}$$

converges, regardless of what x is by using the ratio test. Then the nth term $x^n/n!$ **must** go to 0 as n goes to ∞.

25.1 Summary of Taylor Polynomials and Series

Here are the most important facts about Taylor polynomials and series, all of which were developed in the previous chapter:

1. Of all the polynomials of degree N or less, the one which best approximates the smooth function f for x near a is called the Nth-order Taylor polynomial about $x = a$, and is given by

$$\boxed{P_N(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(N)}(a)}{N!}(x-a)^N}$$

Using sigma notation, this can be written as

$$\boxed{P_N(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n}$$

2. The polynomial P_N has the same derivatives as f at $x = a$, up to and including order N. That is,

$$P_N(a) = f(a), \quad P_N'(a) = f'(a), \quad P_N''(a) = f''(a), \quad P_N^{(3)}(a) = f^{(3)}(a)$$

and so on up to $P_N^{(N)}(a) = f^{(N)}(a)$. The above equations aren't true in general if a is replaced by any other number, or for derivatives of order higher than N. (In fact, the derivatives of P_N of order higher than N are identically 0, since P_N is a polynomial of degree N.)

3. The Nth-order remainder term $R_N(x)$, otherwise known as the Nth-order error term, is simply the difference $f(x) - P_N(x)$. It follows that

$$\boxed{f(x) = P_N(x) + R_N(x)}$$

for any N. The remainder term is given by

$$\boxed{R_N(x) = \frac{f^{(N+1)}(c)}{(N+1)!}(x-a)^{N+1}}$$

where c is some number between x and a which cannot be computed in general

4. So, the complete expression for $f(x)$ is given by

$$\boxed{f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(N+1)}(c)}{(N+1)!}(x-a)^{N+1}}$$

5. The infinite series

$$\boxed{\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n}$$

is called the Taylor series of $f(x)$ about $x = a$. For any particular x, this series may or may not converge. If for any particular x the remainder term $R_N(x)$ converges to 0 as $N \to \infty$, then we can write

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

for that x. That is, $f(x)$ is equal to its Taylor series representation (about $x=a$) at the point x

In the special case where $a=0$, the Taylor series is

$$\sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}x^n$$

This is called the *Maclaurin series* of $f(x)$. So, whenever you see the words "Maclaurin series," you can mentally replace them by "Taylor series about $x=0$."

25.2 Finding Taylor Polynomials and Series

Suppose you want to find a certain Taylor polynomial or series. If you're lucky, you can take a Taylor polynomial or series you already know, manipulate it, and get the polynomial or series you want. Unfortunately, this doesn't always work: sometimes, you need to bust out the formula for the Taylor series of f about $x=a$. This can be a real pain in the butt, however! Differentiating once or twice is bad enough, but differentiating hundreds and thousands of times is ridiculous. Things aren't so bad if you only want to find a Taylor polynomial of low degree, since then you only have to calculate a few derivatives (Page 537).

On the other hand, some functions are really easy to differentiate. One such example is the function f defined by $f(x)=e^x$ (Page 537)

It's really helpful to set up a table of derivatives. In general, the template should look like this:

n	$f^{(n)}(x)$	$f^{(n)}(a)$
0		
1		
2		
3		

For $f(x)=e^x$ about $x=-2$

n	$f^{(n)}(x)$	$f^{(n)}(-2)$
0	e^x	e^{-2}
1	e^x	e^{-2}
2	e^x	e^{-2}
3	e^x	e^{-2}

For $f(x)=\sin(x)$ about $x=\pi/6$

n	$f^{(n)}(x)$	$f^{(n)}(\pi/6)$
0	$\sin(x)$	$1/2$
1	$\cos(x)$	$\sqrt{3}/2$
2	$-\sin(x)$	$-1/2$
3	$-\cos(x)$	$-\sqrt{3}/2$
4	$\sin(x)$	$1/2$

For $f(x) = (1+x)^{1/2}$ about $x = 0$

n	$f^{(n)}(x)$	$f^{(n)}(0)$
0	$(1+x)^{1/2}$	1
1	$\frac{1}{2}(1+x)^{-1/2}$	$1/2$
2	$-\frac{1}{4}(1+x)^{-3/2}$	$-1/4$
3	$\frac{3}{8}(1+x)^{-5/2}$	$3/8$
4	$-\frac{15}{16}(1+x)^{-7/2}$	$-15/16$

Now, let's write down the general formula for the Maclaurin series,

$$f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \frac{f^{(4)}(0)}{4!}x^4 + \cdots$$

$$1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} - \frac{5x^4}{128} + \cdots$$

In fact, it turns out that the remainder term goes to 0 when x is between -1 and 1 (this is tricky to prove!), so we actually have

$$(1+x)^{1/2} = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} - \frac{5x^4}{128} + \cdots$$

when $-1 < x < 1$. This is a special case of the *binomial theorem*, which says that

$$(1+x)^a = 1 + ax + \frac{a(a-1)}{2!}x^2 + \frac{a(a-1)(a-2)}{3!}x^3$$

$$+ \frac{a(a-1)(a-2)(a-3)}{4!}x^4 + \cdots$$

for $-1 < x < 1$. The series on the right-hand side diverges when $x > 1$ or $x < -1$ unless a happens to be a nonnegative integer.

25.3 Estimation Problems Using the Error Term

Consider the following two similar examples:

1. Estimate $e^{1/3}$ using a Taylor polynomial of order 2, and also estimate the error
2. Estimate $e^{1/3}$ with an error no more than $1/10000$

With these two types of problems in mind, check out the general method for solving estimation (or approximation) problems:

1. Look at what you want to estimate, and pick a relevant function f. In our examples above, we want to estimate $e^{\text{something}}$, so set $f(x) = e^x$. Later on, we will set $x = 1/3$, since $f(1/3) = e^{1/3}$, the quantity we want to estimate
2. Pick a number a which is pretty close to this value of x, and so that $f(a)$ is really nice. This means that you should be able to write down $f(a)$ exactly, as well as $f'(a)$, $f''(a)$, and so on. In our example, we'll put $a = 0$, since that's pretty close to $1/3$ and also e^0 is easy to compute

3. Make a table of derivatives of f, just like we did in the previous section. It should have three columns which show the values of n, $f^{(n)}(x)$, and $f^{(n)}(a)$. If you know the order of the Taylor polynomial to use, that's the value of N you'll need; make sure to go up to the $(N+1)$th derivative in the table. Otherwise, just write down as many rows as you can be bothered to; you can always fill in more later if you need to

4. If you don't care about the error in your estimate, skip to step 8. Otherwise, write down the formula for $R_N(x)$:

$$\boxed{R_N(x) = \frac{f^{(N+1)}(c)}{(N+1)!}(x-a)^{N+1}}$$

making sure to write "c is between a and x." As you're writing, replace a by its true value on the fly, including in your comment about c

5. If you know the order of the Taylor polynomial to use, replace N by this number in the above formula. If not, make an educated guess based on how small you need the error to be. The smaller, the higher N should be. For many problems, $N = 2 \text{ or } 3$ will do nicely. If you're wrong, you'll know soon enough; you'll just have to repeat this step and the next two steps with a higher value of N

6. Now, replace x by the value you want in the formula for $R_N(x)$. No unknown variables should be left except for c, and you should write down the possible range of c as an inequality. In our case, with $a = 0$ and $x = 1/3$, we know that c lies in between, so we'd write $0 < c < 1/3$

7. Find the maximum value of $|R_N(x)|$, where c lies in the appropriate interval. This is how big the error can possibly be. If you know the value of N, you're all done with the error estimate. If not, compare the actual error with the one you want. If your actual error is smaller, that's great-you have found a good value of N. Otherwise, you're a little bit screwed-you have to go back to step 5 and try again. (We'll look at some techniques for maximizing $|R_N(x)|$ in Section 25.3.6 below.)

8. Finally, it's time to find the actual estimate! Write down the formula for $P_N(x)$:

$$P_N(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(N)}(a)}{N!}(x-a)^N$$

Now replace a and N by the values from above to get a formula in terms of x alone. Finally, write down the approximation

$$\boxed{f(x) \cong P_N(x)}$$

and plug in the actual value of x that you need. The left-hand side will be the quantity you want, and the right-hand side will be the approximation.

9. One other piece of information is available if you want it: if $R_N(x)$ is positive, your estimate is an underestimate; if $R_N(x)$ is negative, the estimate is an overestimate. These facts follow from the equation

$$\boxed{f(x) = P_N(x) + R_N(x)}$$

- First example

$$e^{1/3}$$

- Second example

$$e^{1/3} \quad \text{error less than } 1/10000$$

- Third example

$$\sqrt{27} \quad \text{error less than } 1/250$$

- Fourth example

$$\sqrt{23} \quad \text{error less than } 1/250$$

- Fifth example

$$\cos(\pi/3 - 0.01)$$

- General techniques for estimating the error term

In all the above examples, we had to estimate the quantity $|f^{(N+1)}(c)|$ for c in some given range. Here are some general tips for doing this:

1. Regardless of the value of c, you can always use the standard inequalities $|\sin(c)| \leq 1$ and $|\cos(c)| \leq 1$
2. If the function $f^{(N+1)}$ is increasing, then its value is biggest at the righthand endpoint
3. If the function $f^{(N+1)}$ is decreasing, then the greatest value of $f^{(N+1)}(c)$ occurs at the left-hand endpoint of the interval
4. In general, you might have to find the critical points of the function $f^{(N+1)}$ in order to maximize it.

25.4 Another Technique for Estimating the Error

Cast your mind back to the alternating series test. The idea is that at each point in the series, adding the next term overshoots the actual value, so the entire error is less than the next term in absolute value (Page 549)

$$|R_N(x)| \leq \left|\frac{f^{(N+1)}(a)}{(N+1)!}(x-a)^{N+1}\right|$$

CHAPTER 26 Taylor and Power Series: How to Solve Problems

In this chapter, we'll look at how to solve four different classes of problems involving Taylor series, Taylor polynomials and power series.

26.1 Convergence of Power Series

Let's say we have a power series about $x = a$:

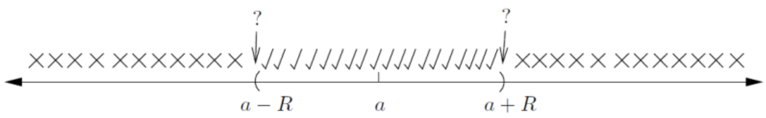
$$\sum_{n=0}^{\infty} a_n (x-a)^n$$

As we saw in the case of geometric series, a power series might converge for some x and diverge for other x. The question that we want to ask is this: given our power series, for which x does it converge, and for which x does it diverge? Furthermore, if the series converges for a specific x, it would be nice to know whether the convergence is absolute or merely conditional.

● Radius of convergence

There are only three possibilities that can occur:

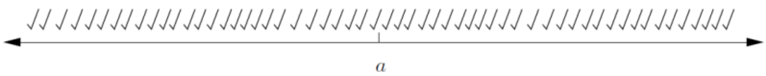
1. There is some number $R > 0$, called the *radius of convergence* of the power series, such that the picture looks like this:



The explanation of this diagram is (Page 553)

An example of this is the geometric series $\sum_{n=0}^{\infty} x^n$. This is a power series with $a = 0$ which converges absolutely when $|x| < 1$ and diverges otherwise. The radius of convergence is therefore equal to 1, and the series diverges at the endpoints 1 and -1

2. The power series might converge absolutely for all x, in which case the diagram looks like this:



In this case, we say that the radius of convergence is ∞. As we saw above, an example of this is the power series for e^x,

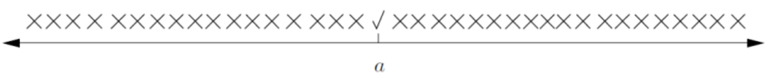
$$\sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Other examples include the Maclaurin series for $\sin(x)$ and $\cos(x)$

3. The power series might converge absolutely only for $x = a$ and diverge for all other x. In this case, the radius of convergence is 0. We'll soon see that this is the case for the series

$$\sum_{n=0}^{\infty} n!\, x^n$$

for example. The picture for this case looks like this:



● How to find the radius and region of convergence

Given a power series, how do we find the radius of convergence? The answer is to use the **ratio test**. Sometimes the root test will be more effective, but for most problems the ratio test is better. Here's the general approach:

1. Write down the limiting absolute ratio; this should always look like

$$\lim_{n\to\infty} \frac{|a_{n+1}(x-a)^{n+1}|}{|a_n(x-a)^n|} = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| |x-a|$$

If instead you use the root test, you should get

$$\lim_{n\to\infty} |a_n(x-a)^n|^{1/n} = \lim_{n\to\infty} |a_n|^{1/n} |x-a|$$

2. Work out the limit. It's important to note that the limit is as $n \to \infty$, not $x \to \infty$. There's a big difference! Regardless of whether you use the ratio test or the root test, the answer should be of the form $L|x-a|$, where L might be a finite number, 0, or even ∞. The important point is that there is a factor of $|x-a|$ present

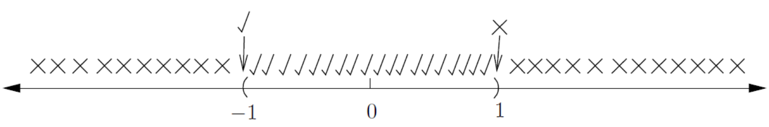
3. In either the ratio test or the root test, the important thing is whether the limit $L|x-a|$ is less than 1, greater than 1, or equal to 1. So, if L is positive, then divide by L to understand everything: if $|x-a| < 1/L$, the power series converges absolutely; if $|x-a| > 1/L$, then the power series diverges; whereas if $|x-a| = 1/L$, then we can't tell and need to check the two endpoints. We are in the first situation from the previous section, and the radius of convergence is $1/L$

4. If $L = 0$, then the limiting ratio is always 0 regardless of the value of x. Since $0 < 1$, this means that the power series converges absolutely for all x, so we are in the second case from the previous section and the radius of convergence is ∞

5. If $L = \infty$, then it looks like the power series never converges. In fact, the series must converge when $x = a$, but it will diverge for every other x and so we are in the third case from the previous section: the radius of convergence is 0

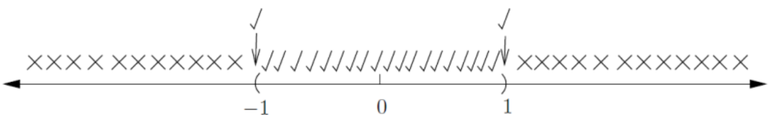
First, consider the power series

$$\sum_{n=2}^{\infty} \frac{x^n}{n\ln(n)}$$



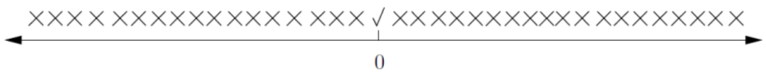
Now consider

$$\sum_{n=2}^{\infty} \frac{x^n}{n(\ln(n))^2}$$



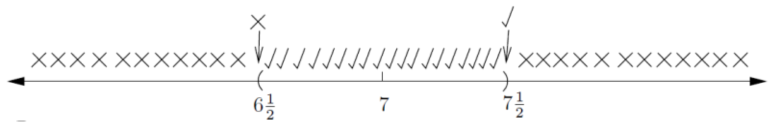
How about

$$\sum_{n=1}^{\infty} n!\, x^n$$



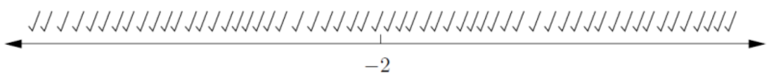
Now consider

$$\sum_{n=1}^{\infty} \frac{(-2)^n}{\sqrt{n}}(x-7)^n$$



Consider the series

$$\sum_{n=1}^{\infty} \frac{3^n}{2^{n^2}}(x+2)^n$$



26.2 Getting New Taylor Series from Old Ones

Let's look at some techniques for **finding** Taylor series. One way to find the Taylor series about $x = a$ of a given function f is to use the formula directly. For most functions, this is a pain. Often a better idea is to use some common Taylor series to synthesize new ones. Of course, you have to know some Taylor series first! It is really useful to have the following five Maclaurin series

1. For $f(x) = e^x$: (in Section 24.2.3)

$$\boxed{e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots}$$

which is true for all real x

2. For $f(x) = \sin(x)$: (in Sections 26.2.2)

$$\boxed{\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots}$$

which is true for all real x

3. For $f(x) = \cos(x)$: (in Section 24.2.3)

$$\boxed{\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots}$$

which is true for all real x

4. For $f(x) = 1/(1-x)$: (in Section 22.2)

$$\boxed{\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots}$$

which is true only for $-1 < x < 1$

5. For $f(x) = \ln(1+x)$ or $f(x) = \ln(1-x)$: (in Sections 26.2.3)

$$\ln(1+x) = \sum_{n=1}^{\infty} -\frac{(-1)^n x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

$$\ln(1-x) = \sum_{n=1}^{\infty} -\frac{x^n}{n} = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots$$

which are **true** for $-1 < x < 1$. (Actually, the first formula is also true for $x = 1$ as well, and the second formula is true for $x = -1$, but this gets a little complicated!)

Anyway, suppose that you've learned all five series. Here's how to manipulate them to get new power series.

● Substitution and Taylor series

The most useful technique is substitution. In a Maclaurin series, you can replace x by a multiple of x^n, where n is an integer, to get a new Maclaurin series. For example, we know that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

for any x; so if you want to find the Maclaurin series for $f(x) = e^{x^2}$, simply replace x by x^2 in the above series to get

$$e^{x^2} = 1 + x^2 + \frac{(x^2)^2}{2!} + \frac{(x^2)^3}{3!} + \frac{(x^2)^4}{4!} + \cdots$$

Let's look at another common example: what is the Maclaurin series for $f(x) = 1/(1+x^2)$?

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots$$

which is valid for $-1 < x < 1$, then replace x by $-x^2$ to get

$$\frac{1}{1+x^2} = \sum_{n=0}^{\infty} (-x^2)^n = \sum_{n=0}^{\infty} (-1)^n x^{2n} = 1 - x^2 + x^4 - x^6 + \cdots$$

which is valid for $-1 < -x^2 < 1$, since the inequalities can reduce to $-1 < x < 1$.

Suppose instead we wanted to work out the Maclaurin series for $1/(1+2x^2)$. Then we would have replaced x by $-2x^2$ instead. This gives

$$\frac{1}{1+2x^2} = \sum_{n=0}^{\infty} (-2x^2)^n = \sum_{n=0}^{\infty} (-1)^n 2^n x^{2n} = 1 - 2x^2 + 4x^4 - 8x^6 + \cdots$$

but this is valid only for $-1 < -2x^2 < 1$, which can reduce to $-1/\sqrt{2} < x < 1/\sqrt{2}$.

Now, suppose you start with the following equation, which is true for all real x:

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

If you replace x by $(x-18)$, you get a Taylor series about $x = 18$ instead:

$$\sin(x-18) = (x-18) - \frac{(x-18)^3}{3!} + \frac{(x-18)^5}{5!} - \frac{(x-18)^7}{7!} + \cdots$$

The right-hand side is not the Taylor series about $x = 18$ of $\sin(x)$, because the left-hand side is no longer $\sin(x)$-it's $\sin(x-18)$. It's the Taylor series about $x = 18$ of $\sin(x-18)$.
The moral of this last example is that if you replace x by $(x-a)$, then you get a Taylor series about $x = a$ instead of a Maclaurin series, but the function is different. This can still be useful. For example, to find the Taylor series of $\ln(x)$ about $x = 1$, start with one of the formulas from the previous section:

$$\ln(1+x) = \sum_{n=1}^{\infty} -\frac{(-1)^n x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \quad \text{for} -1 < x < 1$$

Now, let's replace x by $(x-1)$

$$\ln(x) = \sum_{n=1}^{\infty} -\frac{(-1)^n (x-1)^n}{n} = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \cdots$$

$$\text{for } -1 < (x-1) < 1 \text{ or } 0 < x < 2$$

By the way, the substitution technique can also be used to find Taylor polynomials, but you have to be careful to get the order right. For example, if you take $f(x) = e^x$ and $a = 0$, the Taylor polynomial of order 3 is

$$P_3(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}$$

Now if $f(x) = e^{x^2}$, it's a mistake to replace x by x^2 in the above polynomial and claim the third-order Taylor polynomial of g is

$$P_3(x) = 1 + x^2 + \frac{x^4}{2!} + \frac{x^6}{3!}$$

This is actually the **sixth-order** Taylor polynomial of g about 0, so the left-hand side should say $P_6(x)$ instead of $P_3(x)$.

● Differentiating Taylor series

If a power series converges to a differentiable function f on an open interval (a, b), then it turns out that you can differentiate the series term-by-term to get a new series which converges to $f'(x)$ on the same interval. The situation at the endpoints a and b is a little trickier: the differentiated series might diverge even if the original series converges. So check the endpoints separately.
Our first example is to find the Maclaurin series for $\sin(x)$, assuming that we know the Maclaurin series for $\cos(x)$ is given by

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \cdots$$

the formula is valid for all x. If you differentiate both sides, term-by-term on the right, you can get

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

Here's another example of differentiating a power series. Suppose you want to find the Maclaurin series for $f(x) = 1/(1+x)^2$. The best way would be to start with the series for $1/(1+x)$, which is obtained from the standard geometric series (#4 above) by replacing x by $-x$

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - \cdots$$

this is valid for $-1 < x < 1$. Then differentiate both sides, term-by-term on the right-hand side, you can get

$$\frac{1}{(1+x)^2} = 1 - 2x + 3x^2 - 4x^3 + \cdots = \sum_{n=0}^{\infty} (-1)^n (n+1) x^n$$

This is valid for $-1 < x < 1$.

Once again, you can apply these ideas to Taylor polynomials; you just have to be careful with orders, once again. Since differentiating a polynomial knocks the degree down by one, the differentiated Taylor polynomial is order one less than the original polynomial.

● Integrating Taylor series

You can also integrate a power series term-by-term. The new series converges in the same interval as the old one (except perhaps at the endpoints of the interval of convergence). If you use an indefinite integral, don't forget the constant! Let's see a few examples. First, let's try to prove the following formula for $\ln(1-x)$

$$\ln(1-x) = \sum_{n=1}^{\infty} -\frac{x^n}{n} = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots$$

For $-1 < x < 1$. To do it, we'll use the geometric series formula, which is #4 in Section 26.2:

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots$$

valid for $-1 < x < 1$

$$\int \frac{1}{1-x} dx = \int \sum_{n=0}^{\infty} x^n \, dx = \int (1 + x + x^2 + x^3 + \cdots) dx$$

$$-\ln(1-x) = C + \sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1} = C + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots$$

It's a good idea to put the constant first instead of as $+C$ at the end, since it's really the zeroth-dgree term in the power series. Now we have to find out what C actually is. The best way is to substitute $x = 0$

$$-\ln(1-0) = C + 0 + \frac{0^2}{2} + \frac{0^3}{3} + \frac{0^4}{4} + \cdots$$

which reduces to $C = 0$. So you get

$$\ln(1-x) = \sum_{n=0}^{\infty} -\frac{x^n}{n} = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots$$

valid for $-1 < x < 1$.

Another example: how would you find the Maclaurin series for $\tan^{-1}(x)$? We can be really sneaky and integrate a series we already know. Let's see, $\tan^{-1}(x)$ is an antiderivative of $1/(1+x^2)$

$$\int \frac{1}{1+x^2} dx = \int (1 - x^2 + x^4 - x^6 + \cdots) \, dx$$

$$\tan^{-1}(x) = C + x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots$$

Now we substitute $x = 0$ to find out $C = 0$. Since the original series for $1/(1+x^2)$ converges when $-1 < x < 1$, so does the series for $\tan^{-1}(x)$.

Let's look at an example of a definite integral. Suppose that a function f is defined by

$$f(x) = \int_0^x \sin(t^3)\, dt$$

Start by finding the series for $\sin(t^3)$

$$\sin(t^3) = t^3 - \frac{(t^3)^3}{3!} + \frac{(t^3)^5}{5!} - \frac{(t^3)^7}{7!} + \cdots$$

$$f(x) = \int_0^x \sin(t^3)\, dt = \int_0^x \left(t^3 - \frac{t^9}{3!} + \frac{t^{15}}{5!} - \frac{t^{21}}{7!} + \cdots\right) dt$$

$$f(x) = \left(\frac{t^4}{4} - \frac{t^{10}}{10 \cdot 3!} + \frac{t^{16}}{16 \cdot 5!} - \frac{t^{22}}{22 \cdot 7!} + \cdots\right)\Big|_0^x$$

this is valid for all real x.

You can also apply the above integration techniques to Taylor polynomials; this time the order of the Taylor polynomial increases by 1

● Adding and subtracting Taylor series

If you know the Taylor series about $x = a$ for two functions f and g, then the Taylor series for the sum $f(x) + g(x)$ is of course the sum of the two respective Taylor series, at least in the overlap of the regions where the Taylor series converge. The same goes for the difference $f(x) - g(x)$. The only thing you need to do in practice is to group terms of the same degree together, and worry about where the resulting series converges (Problem: Ignore higher order in Page 565).

● Multiplying Taylor series

You can also multiply two Taylor series to get a new one which converges to the product of the two relevant functions, at least in the intersection of the regions where the Taylor series converge. Writing this in sigma notation can get pretty messy and usually involves double sums. Normally one is interested in the first few terms of a series (Problem: Ignore higher order in Page 566).

● Dividing Taylor series

You can do exactly the same thing with quotients by using long division. The trick is to ignore all but the terms of order up to the one you are interested in. For example, to find the Maclaurin series for $f(x) = \sec(x)$ up to fourth order, first write $\sec(x)$ as $1/\cos(x)$, then set up a long division just as you do with polynomials

$$\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{24} - \cdots$$

in the long division for $1/\cos(x)$:

$$
\begin{array}{r|l}
 & 1 + \frac{1}{2}x^2 + \frac{5}{24}x^4 + \cdots \\
\hline
1 + 0x - \frac{1}{2}x^2 + 0x^3 + \frac{1}{24}x^4 - \cdots & 1 + 0x + 0x^2 + 0x^3 + 0x^4 + \cdots \\
 & 1 + 0x - \frac{1}{2}x^2 + 0x^3 + \frac{1}{24}x^4 + \cdots \\
\cline{2-2}
 & \frac{1}{2}x^2 + 0x^3 - \frac{1}{24}x^4 + \cdots \\
 & \frac{1}{2}x^2 + 0x^3 - \frac{1}{4}x^4 + \cdots \\
\cline{2-2}
 & \frac{5}{24}x^4 + \cdots
\end{array}
$$

So the Maclaurin series for $\sec(x)$ is $1 + x^2/2 + 5x^4/24 + \cdots$, up to terms of fourth order.

So the moral of the story is that you may not have to differentiate over and over again and use the formula for Taylor series. If you're lucky, you can instead use some of the five basic series, plus one or more of the techniques of substitution, differentiation, integration, addition, subtraction, multiplication, and division.

26.3 Using Power and Taylor Series to Find Derivatives

Recall the formula for the nth coefficient of the Taylor series of $f(x)$ about $x = a$:

$$a_n = \frac{f^{(n)}(a)}{n!}$$

Let's multiply through by $n!$ to arrive at the following formula:

$$\boxed{f^{(n)}(a) = n! \times a_n}$$

In words, this means that

$$\boxed{f^{(n)}(a) = n! \times \begin{pmatrix} \text{the coefficient of } (x-a)^n \text{ in the} \\ \text{Taylor series of } f(x) \text{ about } x = a \end{pmatrix}}$$

So if you know the Taylor series of a function about some point a, you can easily find the derivatives of that function at a. This is all you get! There's no information about the value of the derivatives at any other value of x; it's only $x = a$.

26.4 Using Maclaurin Series to Find Limits

You can also use some Taylor series to find certain limits. In particular, if you have a limit like

$$\lim_{x \to 0} \frac{f(x)}{g(x)}$$

where both the numerator and the denominator are 0 when $x = 0$, then you could use l'Hôpital's Rule (Page 570): The correct method is to replace everything in sight by enough terms of the appropriate Maclaurin series. Basically, if everything cancels out, you haven't used enough terms, whereas if something is still left, you've gone far enough and can proceed. So, it's better to use more terms rather than fewer.

Here's the real reason all the above limits work: if f has a Maclaurin series with lowest-degree term $a_N x^N$, then

$$f(x) \sim a_N x^N \quad \text{as } x \to 0$$

CHAPTER 27 Parametric Equations and Polar Coordinates

So far, we've sketched the graphs of many equations of the form $y = f(x)$ with respect to Cartesian coordinates. Now we're going to look at things in a different way.

27.1 Parametric Equations

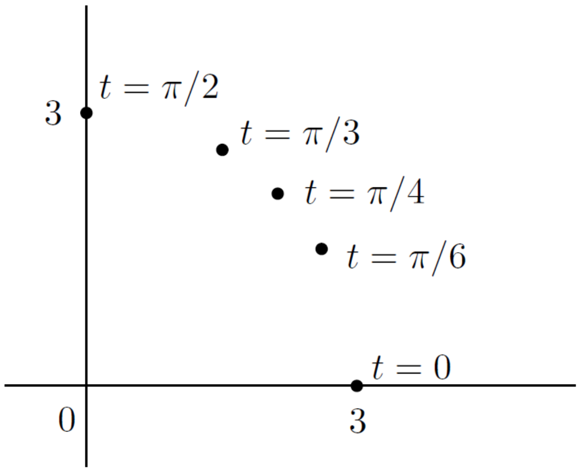
When you write an equation like $y = x^2 \sin(x)$, you are expressing y as a function of x. So if you have a particular value of x in mind, then you can easily find the corresponding value of y by plugging that value of x into the above equation. On the other hand, consider the relation $x^2 + y^2 = 9$, it's a little harder and you can write $y = \pm\sqrt{9 - x^2}$.

Now let's try a different approach: suppose that both x and y are functions of another variable t. For example, we could set

$$x = 3\cos(t) \quad \text{and} \quad y = 3\sin(t)$$

If you like, you could even write $x(t) = 3\cos(t)$ to emphasize the x as a function of t. Then you can get corresponding values for **both** x and y by plugging your value of t into the above equations. The variable t is called a *parameter*, and the above equations are called *parametric equations*

t	0	$\pi/6$	$\pi/4$	$\pi/3$	$\pi/2$
x	3	$3\sqrt{3}/2$	$3/\sqrt{2}$	$3/2$	0
y	0	$3/2$	$3/\sqrt{2}$	$3\sqrt{3}/2$	3



Notice that if you pick a point (x, y) on the circle, there isn't just one value of t which corresponds to that point! There are infinitely many, all separated by multiples of 2π.

So, the above pair of parametric equations describes the circle $x^2 + y^2 = 9$, at least if you let t range over a large enough interval-for example, $[0, 2\pi)$. You can say that

$$x = 3\cos(t) \quad \text{and} \quad y = 3\sin(t), \quad \text{where } 0 \le t \le 2\pi$$

is a *parametrization* of $x^2 + y^2 = 9$. Is the graph of $x^2 + y^2 = 9$ the same as the graph of the above parametrization? No, the parametric version tells you a little more; it allows you to find the extra information of direction and speed.

There are many other ways to draw the same circle $x^2 + y^2 = 9$. For example, $x = 3\cos(2t)$ and $y = 3\sin(2t)$, is twice as fast as previous one. Alternatively, $x = 3\sin(t)$ and $y = 3\cos(t)$, starts at $(0, 3)$ and go clockwise around the circle instead of counterclockwise.

How would you find a parametrization for $x^2 + 4y^2 = 9$? (Page 577)

$$x = 3\cos(t) \quad \text{and} \quad y = \frac{3}{2}\sin(t), \quad \text{where } 0 \le t \le 2\pi$$

How about $x^6 + y^6 = 64$?

$$x^6 + y^6 = \left(2\cos^{1/3}(t)\right)^6 + \left(2\sin^{1/3}(t)\right)^6 = 64\cos^2(t) + 64\sin^2(t) = 64$$

● Derivatives of parametric equations

This is a calculus book, so we'd better do some calculus with this parametric stuff. To find the equation of a tangent line to the **curve**, we'll need a derivative, of course. Since x and y are both functions of t, we have to use the chain rule. This says that

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}$$

now divide through by dx/dt and rearrange to get

$$\boxed{\frac{dy}{dx} = \frac{dy/dt}{dx/dt}}$$

If you are thinking of x as $x(t)$ and similarly for y, then you can rewrite this equation as

$$\frac{dy}{dx} = \frac{y'(t)}{x'(t)}$$

Let's look at three examples of how to use this.

First, suppose that we want the slope and equation of the tangent line at the point corresponding to $t = 1/2$ on the parametric curve defined by

$$x = e^{-2t}, \quad y = \sin^{-1}(t), \quad -1 < t < 1$$

Differentiating, we find that

$$\frac{dx}{dt} = -2e^{-2t} \quad \text{and} \quad \frac{dy}{dt} = \frac{1}{\sqrt{1-t^2}}$$

Since we only care about the point $t = 1/2$

$$\frac{dx}{dt} = -2e^{-1} = -\frac{2}{e} \quad \text{and} \quad \frac{dy}{dt} = \frac{1}{\sqrt{1-1/4}} = \frac{2}{\sqrt{3}}$$

So at $t = 1/2$, we have

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{2/\sqrt{3}}{-2/e} = -\frac{e}{\sqrt{3}}$$

Well, then we have to put $t = 1/2$ in the original equations for x and y above to see that $x = e^{-2\cdot(1/2)} = 1/e$ and $y = \sin^{-1}(1/2) = \pi/6$. So the equation of tangent line is

$$y - \frac{\pi}{6} = -\frac{e}{\sqrt{3}}\left(x - \frac{1}{e}\right)$$

Now for a trickier example. Let's try using our parametrization $x = 2\cos^{1/3}(t)$ and $y = 2\sin^{1/3}(t)$ from the curve $x^6 + y^6 = 64$ at point $\left(-2^{5/6},\ 2^{5/6}\right)$; here $0 \le t \le 2\pi$. We have

$$\frac{dx}{dt} = -\frac{2}{3}\cos^{-2/3}(t)\sin(t) \quad \text{and} \quad \frac{dy}{dt} = \frac{2}{3}\sin^{-2/3}(t)\cos(t)$$

So by the chain rule,

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{\frac{2}{3}\sin^{-2/3}(t)\cos(t)}{-\frac{2}{3}\cos^{-2/3}(t)\sin(t)} = -\frac{\cos^{5/3}(t)}{\sin^{5/3}(t)}$$

We can see that

$$x = 2\cos^{1/3}(t) = -2^{\frac{5}{6}}, \text{so } \cos(t) = -1/\sqrt{2}$$

$$y = 2\sin^{1/3}(t) = 2^{5/6}, \text{so } \sin(t) = 1/\sqrt{2}$$

$$\frac{dy}{dx} = -\frac{\cos^{5/3}(t)}{\sin^{5/3}(t)} = -\frac{\left(-1/\sqrt{2}\right)^{5/3}}{\left(1/\sqrt{2}\right)^{5/3}} = 1$$

So the tangent line is

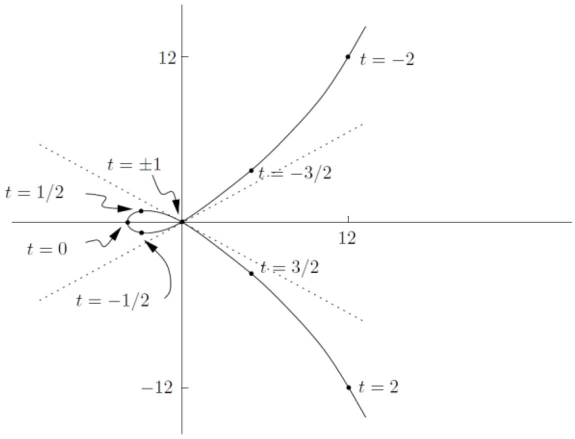
$$y - 2^{5/6} = 1\left(x - \left(-2^{5/6}\right)\right)$$

Now for our trickiest example (conceptually speaking, at least). Suppose that we are given the following parametric equations:

$$x = 4t^2 - 4 \quad \text{and} \quad y = 2t - 2t^3 \quad \text{for all real } t$$

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{2 - 6t^2}{8t} = \frac{1}{4t} - \frac{3t}{4}$$

t	-2	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2
x	12	5	0	-3	-4	-3	0	5	12
y	12	$\frac{15}{4}$	0	$-\frac{3}{4}$	0	$\frac{3}{4}$	0	$-\frac{15}{4}$	-12



Suppose that now we want to find the **second** derivative of the above parametric equations at $t = 1$. The secret to finding d^2y/dx^2 is to consider it as dy'/dx. That is, think of the second derivative as the derivative of y', which itself is the derivative of y with respect to x.

We already saw above that

$$y' = \frac{dy}{dx} = \frac{1}{4t} - \frac{3t}{4}$$

We now use the chain rule to write

$$\frac{d^2y}{dx^2} = \frac{dy'}{dx} = \frac{dy'/dt}{dx/dt} = \frac{\frac{d}{dt}\left(\frac{1}{4t} - \frac{3t}{4}\right)}{\frac{d}{dt}(4t^2 - 4)} = \frac{\frac{-1}{4t^2} - \frac{3}{4}}{8t} = -\frac{1}{32t^3} - \frac{3}{32t}$$

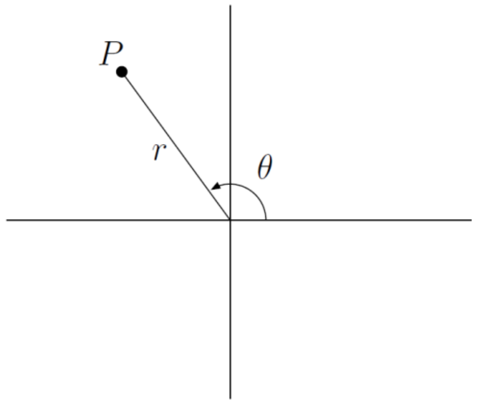
Now we can finally substitute $t = 1$

$$\frac{d^2y}{dx^2} = -\frac{1}{32} - \frac{3}{32} = -\frac{1}{8}$$

27.2 Polar Coordinates

- Converting to and from polar coordinates

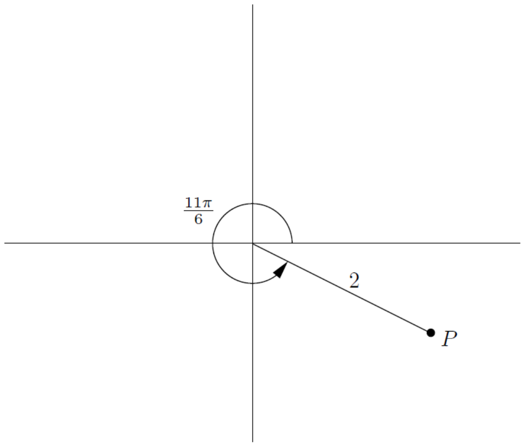
Consider the point (r, θ) in polar coordinates, which could look something like this:



Remember, your friend started at the origin facing toward the positive direction on the x-axis, then turned counterclockwise an angle θ, then marched forward r units to get to the point P. What are the Cartesian coordinates (x, y) of P? Well, we know that $\cos(\theta) = x/r$ and $\sin(\theta) = y/r$, so that gives us

$$\boxed{x = r\cos(\theta) \text{ and } y = r\sin(\theta)}$$

(Compare this with the example $x = 3\cos(t)$, $y = 3\sin(t)$ from Section 27.1 above.) Anyway, these equations show us how to convert from polar to Cartesian coordinates. For example, given in polar coordinates by $(2, 11\pi/6)$, it's not a bad idea to draw a picture:

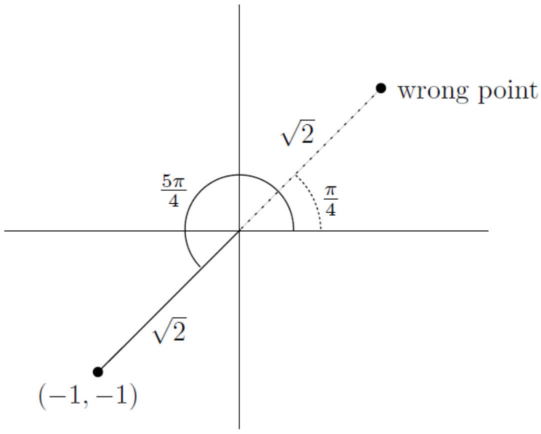


The Cartesian coordinates are $(\sqrt{3}, -1)$.

It's always easier translating from a foreign language into your native language than the other way around; the same thing happens with polar coordinates. It's a little harder getting from Cartesian coordinates to polar coordinates. Here's a summary of the situation:

$$\boxed{r^2 = x^2 + y^2 \text{ and } \tan(\theta) = \frac{y}{x} \quad \text{if } x \neq 0\text{, but check the quadrant!}}$$

Let's look at an example: suppose we want to write $(-1, -1)$ in polar coordinates

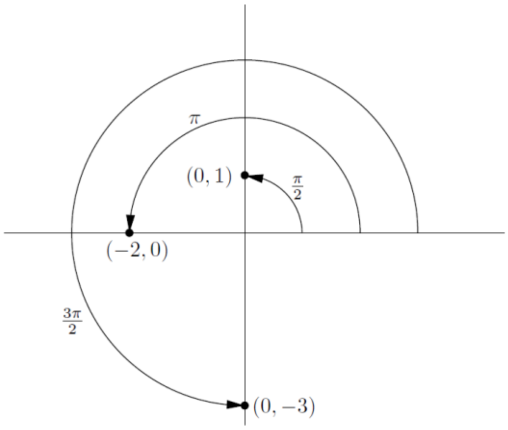


We now have two ways of writing $(-1,-1)$ in polar coordinates: $(\sqrt{2}, 5\pi/4)$ and $(-\sqrt{2}, \pi/4)$. The complete list of points in polar coordinates we could use is as follows:

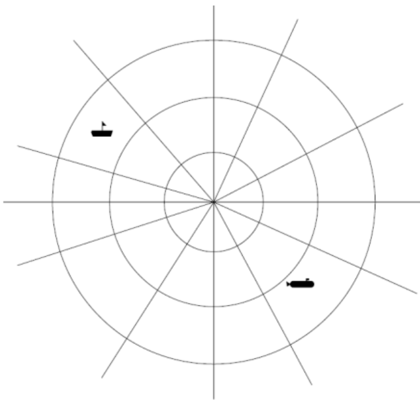
$$\left(\sqrt{2}, \frac{5\pi}{4} + 2\pi n\right), \quad \left(-\sqrt{2}, \frac{\pi}{4} + 2\pi n\right) \quad \text{where } n \text{ is an integer}$$

The convention is usually to choose the one where $r \geq 0$ and θ lies between 0 and 2π.

A few more examples: what are polar coordinates for the points with Cartesian coordinates $(0,1)$, $(-2,0)$, and $(0,-3)$? Let's plot these points on the same set of axes:



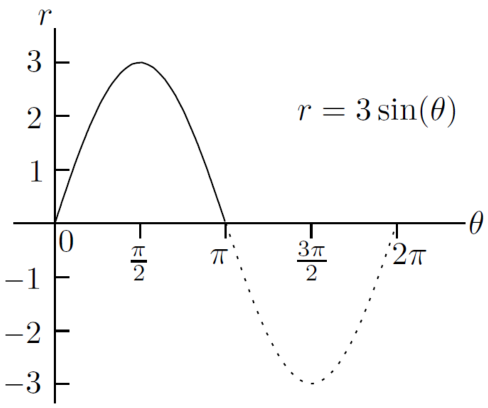
You can get in some trouble using the formula $\tan(\theta) = y/x$ from above. Just look at the picture: $(0,1)$ has polar coordinates $(1, \pi/2)$, $(-2,0)$ has polar coordinates $(2, \pi)$, and $(0,-3)$ has polar coordinates $(3, 3\pi/2)$.



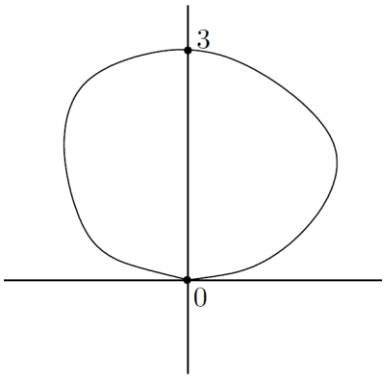
This is just a "grid" in polar coordinates

● Sketching curves in polar coordinates

Suppose you know that $r = f(\theta)$ for some function f, and you want to sketch the graph of all points (r, θ) in polar coordinates where $r = f(\theta)$ for θ in some given range. This isn't so easy to do. Probably the best way to proceed is to draw up a table of values and plot points. It can also be helpful to sketch $r = f(\theta)$ in Cartesian coordinates first. For example, to sketch $r = 3\sin(\theta)$ in polar coordinates, where $0 \le \theta \le \pi$:



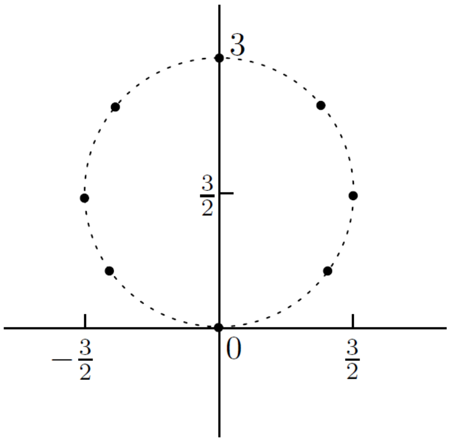
So the curve we want looks something like this:



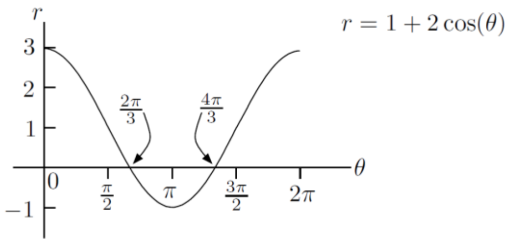
This looks a little pathetic. To tighten it up, we can write down the following table of values:

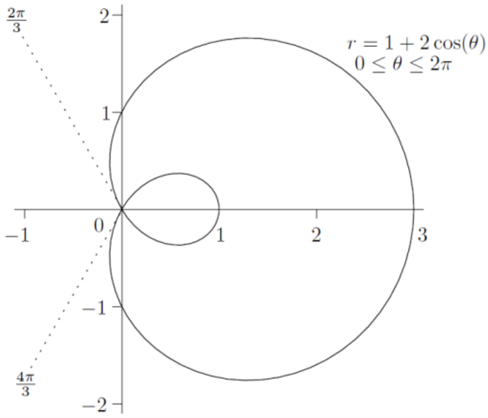
θ	0	$\pi/6$	$\pi/4$	$\pi/3$	$\pi/2$	$2\pi/3$	$3\pi/4$	$5\pi/6$	π
r	0	$3/2$	$3/\sqrt{2}$	$3\sqrt{3}/2$	3	$3\sqrt{3}/2$	$3/\sqrt{2}$	$3/2$	0

Plotting these points leads to the following picture:

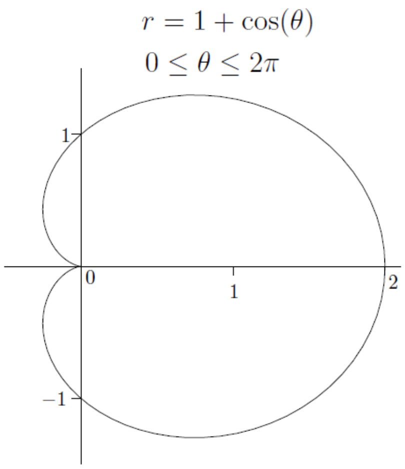


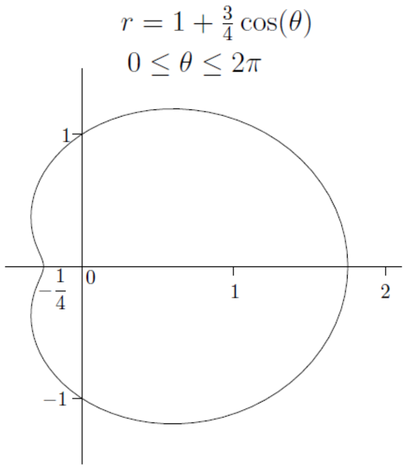
Let's look at another example. Suppose that we want to sketch the curve $r = 1 + 2\cos(\theta)$, where $0 \le \theta \le 2\pi$

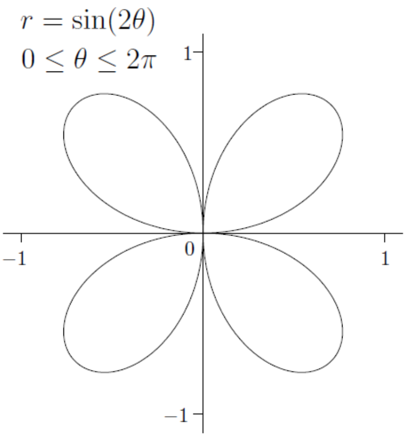


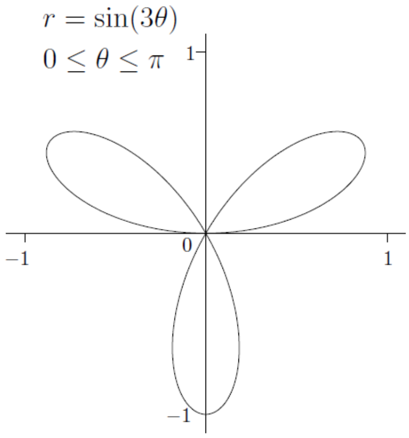


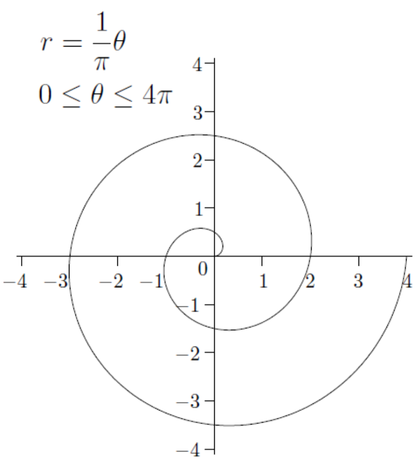
You should try sketching a lot of polar curves until you feel you're going round in circles

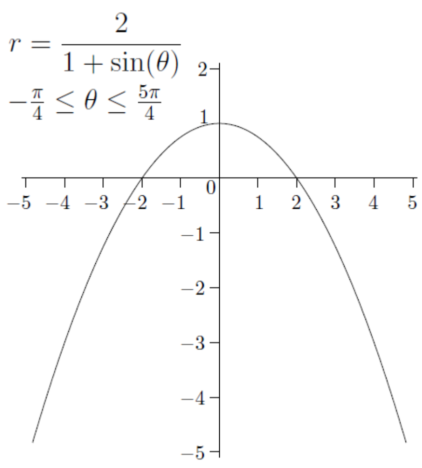












Some facts about the above curves:

1. The curve given by $r = 1 + \cos(\theta)$ is called a *cardioid*. The curve $r = 1 + \frac{3}{4}\cos(\theta)$ is an example of a limacon, of which the cardioid is a special case
2. In the above graph of $r = \sin(3\theta)$, the angle θ only goes from 0 to π. As θ goes from π to 2π, the graph is retraced, just as in the case of the circle $r = \sin(\theta)$
3. The curve given by $r = \theta/\pi$ is an example of a *spiral of Archimedes*. This is not periodic: as θ increases, the spiral gets bigger and bigger
4. The curve given by $r = 2/(1 + \sin(\theta))$ looks like a parabola. In fact, you should try to show that the above equation becomes $x^2 = 4 - 4y$ in Cartesian coordinates

● Finding tangents to polar curves

Luckily, finding tangents to polar curves is just a special case of finding tangents to curves given by parametric equations. Let's see how it works in the case of polar coordinates.

We have $r = f(\theta)$, and we'd like to find the tangent to the curve at some point on the curve. Using $x = r\cos(\theta)$ and $y = r\sin(\theta)$, we can write

$$x = f(\theta)\cos(\theta) \quad \text{and} \quad y = f(\theta)\sin(\theta)$$

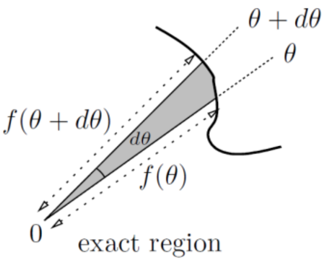
this means that x and y are parametrized by θ. By the formula from Section 27.1.1 above, we have

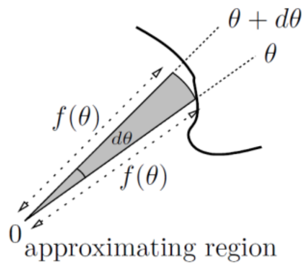
$$\frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta}$$

This gives the slope of the tangent in general. Finally, we just have to plug in the value of θ we care about (Page 590).

● Finding areas enclosed by polar curves

If we want to find the area enclosed by the polar curve $r = f(\theta)$, where f is assumed to be continuous, then we're going to have to integrate something. But what? We just have to set up the correct Riemann sum. Suppose we take a small chunk of angle between θ and $\theta + d\theta$. Then as we move counterclockwise along this chunk of angle, r meanders from $f(\theta)$ to $f(\theta + d\theta)$. If $d\theta$ is very small, then r doesn't have a chance to move far away from $f(\theta)$, so we can approximate the wedge we're looking for by a thin slice of pie of radius $r = f(\theta)$ units and angle $d\theta$, centered at the origin, as shown in the following diagram:



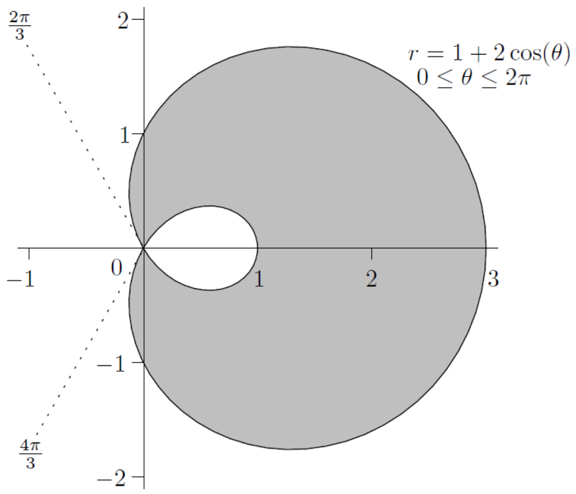


The area of a sector is one half of the radius squared, multiplied by the angle of the sector (in radians $\left(\frac{\theta}{2\pi}, \text{and the area of a circle is } \pi r^2\right)$, of course!). So, we can approximate the area of the wedge (in square units) by $\frac{1}{2}\left(f(\theta)\right)^2 d\theta$, which is just $\frac{1}{2}r^2 d\theta$. The total area

$$\boxed{(\text{area inside } r = f(\theta) \text{ between } \theta = \theta_0 \text{ and } \theta = \theta_1) = \int_{\theta_0}^{\theta_1} \frac{1}{2} r^2 d\theta}$$

As usual, the area is given in square units.

Example (Page 592)



CHAPTER 28 Complex Numbers

28.1 The Basics

It kind of sucks that you can't take the square root of -1. So, we'll just do it anyway. Let's just create a square root of -1 and call it i. OK, so then we must have $i^2 = -1$. Is i the only square root of -1? No, $-i$ should also be a square root, since if there were any justice in the world, then

$$(-i)^2 = (-1)^2(i)^2 = 1(-1) = -1$$

Since $i^2 + 1 = 0$ and $(-i)^2 + 1 = 0$, we now have two roots for the quadratic $x^2 + 1$ after all-but they are not real: they are imaginary. How about $2i$? That's also imaginary. In fact, $(2i)^2 = 2^2 i^2 = 4(-1) = -4$, so $(2i)^2$ is a negative number. So, when we say that a number is *imaginary*, we mean that its square is a negative number. The only imaginary numbers are of the form yi where y is a real number not equal to 0. You can also write iy instead of yi.

Now, you can add or subtract real and imaginary numbers, for example $2 - 3i$, but you can't simplify the result. In this way, we get all the complex numbers, which are all the numbers of the form $x + iy$, where x and y are real. The set of all complex numbers is normally denoted by the symbol C. Notice that all imaginary numbers are complex numbers ($2i = 0 + 2i$). All real numbers are also complex numbers ($-13 = -13 + 0i$). Every complex number has a real and an imaginary part. If $z = x + iy$, then the real part is x and the imaginary part is y. These are written as $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$, respectively. For example, $\mathrm{Re}(2 - 3i) = 2$ and $\mathrm{Im}(2 - 3i) = -3$. Note that $\mathrm{Im}(2 - 3i)$ is not $-3i$, it's just -3.

Adding and subtracting complex numbers is pretty easy. Just add (or subtract) the real parts, and then do the imaginary parts (Page 596).

Multiplication isn't much harder-you just expand, but remember to change i^2 into -1 whenever you see it (Page 596).

By the way, what is i^3? How about i^4? i^5? $i^3 = -i$, $i^4 = 1$, $i^5 = i$. In fact, because $i^4 = 1$, we can see that the powers of i keep on cycling through $1, i, -1, -i$.

How about division? That's a little trickier, but not much. The technique is very similar to rationalizing the denominator. It's inspired by the following observation: If you have a complex number $x + iy$ and multiply it by the complex number $x - iy$, you get a real number

$$(x + iy)(x - iy) = x^2 - (iy)^2 = x^2 - i^2 y^2 = x^2 + y^2$$

If $z = x + iy$, the related number $x - iy$ is so important that it has a name: it is called the *complex conjugate* of $x + iy$ and denoted $\bar{z}$. Note that the complex conjugate of a real number is the same number. Now as the above formula shows, a number multiplied by its complex conjugate is real. Inspired by Pythagoras' Theorem and the above formula, given a complex number $z = x + iy$, let's define the *modulus* of z to be $\sqrt{x^2 + y^2}$. We write the modulus of z as $|z|$. So

$$|x + iy| = \sqrt{x^2 + y^2}$$

It's exactly the same as the absolute value. Our notation for modulus is completely consistent with the previous notation for absolute value. In fact, think of the modulus as a beefed-up version of absolute value. Anyway, the difference of two squares formula above shows that a complex number multiplied by its complex conjugate is the square of its modulus. That is,

$$z\bar{z} = |z|^2$$

Now let's see how to solve quadratic equations. For example, let's say that you want to solve $x^2 + 3x + 14 = 0$. Just use the quadratic formula and the fact that $\sqrt{-1} = \pm i$ to write

$$x = \frac{-3 \pm \sqrt{3^2 - 4 \times 1 \times 14}}{2} = \frac{-3 \pm \sqrt{-47}}{2} = -\frac{3}{2} \pm \frac{\sqrt{47}}{2} i$$

Notice that we have simplified $\pm\sqrt{-47}$ as $\pm\sqrt{47}i$. Now, how about if you have a quadratic whose coefficients are complex numbers? The quadratic formula still works, but you may well have to take the square root of a complex number, not just a negative number.

● Complex exponentials

We've discussed how to add and multiply complex numbers. How about exponentiating them? (Page 598)

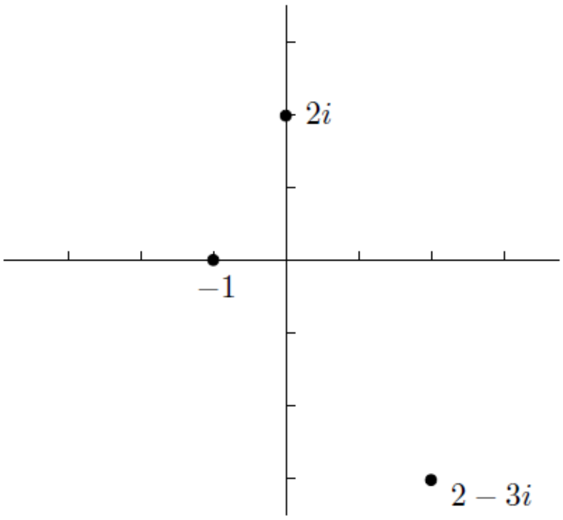
We'll define e^z, for any complex number z, by the following equation:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

And we can show that $e^z e^w = e^{z+w}$.

28.2 The Complex Plane

Real numbers are usually represented as points on a number line, which is one-dimensional. Complex numbers literally have an extra dimension. Indeed, if $z = x + iy$, we can't squish all the information into just one real number. Instead of a real number line, we'll use a complex number plane. The complex number $z = x + iy$ will be represented as the point (x, y) in Cartesian coordinates. It's pretty easy to plot complex number like $2 - 3i$, $2i$, and -1:



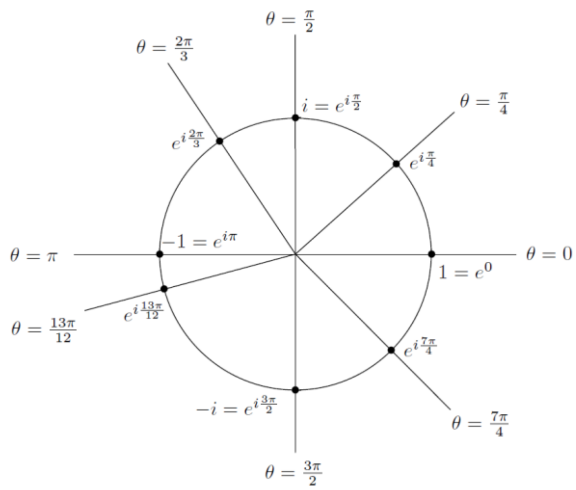
You should think of each point as representing one complex number, rather than as a pair of real numbers.

In the previous chapter, we saw that you can also express every point in the plane in polar coordinates instead. So suppose you have a point in the complex plane which has polar coordinates (r, θ). What is the complex number represented by that point? Well, we can convert to Cartesian coordinates using $x = r\cos(\theta)$ and $y = r\sin(\theta)$. So the point (r, θ) in polar coordinates represents the complex number $z = x + iy = r\cos(\theta) + ir\sin(\theta)$. In particular, if $r = 1$, then z is just $\cos(\theta) + i\sin(\theta)$.

Now, there's a pretty bizarre and funky identity, due to Euler, which is really important:

$$\boxed{e^{i\theta} = \cos(\theta) + i\sin(\theta)}$$

This is true for all real θ. This means that the complex number $e^{i\theta}$, has polar coordinates $(1, \theta)$ when you plot it on the complex number plane. So $e^{i\theta}$ lives on the unit circle and has angle θ from the positive x-axis. The following picture shows a few positions of $e^{i\theta}$ for different values of θ:



For points not on the unit circle, you just have to multiply by r. Specifically, we saw that if z is represented by the point (r, θ) in polar coordinates, then $z = r\cos(\theta) + ir\sin(\theta)$. By Euler's identity, this means that $z = re^{i\theta}$. So we have shown that

$$\boxed{\text{if } (x, y) \text{ and } (r, \theta) \text{ are the same point, then } x + iy = re^{i\theta}}$$

Let's say that a complex number like $re^{i\theta}$ is in *polar form*, (as opposed to $x + iy$, which is in *Cartesian form*). This formula looks a bit strange, but it's true enough (Page 600).

Let's agree that when we're dealing with complex numbers, we'll never let r be negative (Page 600).

$e^{i\theta}$ is **periodic** in θ with period 2π.

● Converting to and from polar form

To convert a complex number from polar to Cartesian form, just use Euler's identity directly (that's $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ in case you have already forgotten!). For example

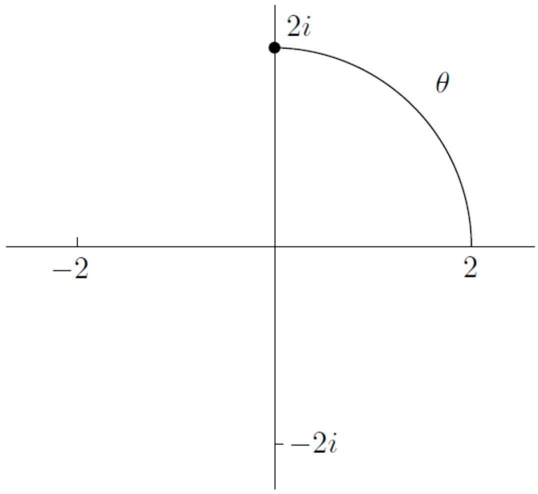
$$2e^{i(5\pi/6)} = 2\left(\cos\left(\frac{5\pi}{6}\right) + i\sin\left(\frac{5\pi}{6}\right)\right) = 2\left(-\frac{\sqrt{3}}{2} + i\frac{1}{2}\right) = -\sqrt{3} + i$$

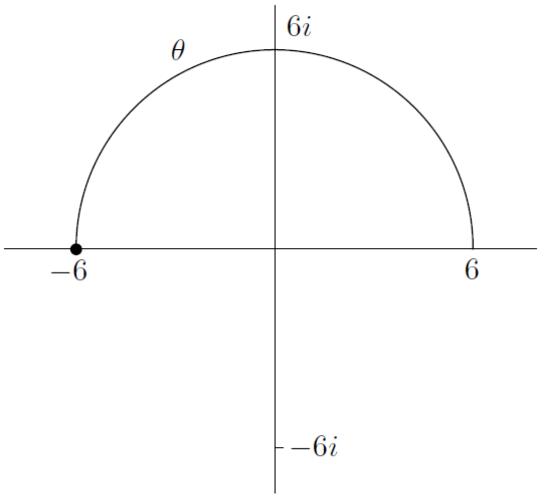
On the other hand, converting from Cartesian to polar form is more difficult, as we observed in Section 27.2.1. There we saw that

$$r = \sqrt{x^2 + y^2} \quad \text{and} \quad \tan(\theta) = \frac{y}{x}$$

where we have now dropped the possible solution $r = -\sqrt{x^2 + y^2}$ since we want $r \geq 0$ for complex numbers. By the way, we defined the modulus of z to be $|z| = \sqrt{x^2 + y^2}$. So r is the same as $|z|$. The modulus $|z|$ is therefore the distance from the point z to the origin (in the complex number plane). The angle θ is called the *argument* of z and is written $\arg(z)$. (Normally one requires that $0 \leq \arg(z) < 2\pi$ so that there's no ambiguity.)

Let's revisit a couple of examples that might seem confusing (Page 602).





28.3 Taking Large Powers of Complex Numbers

Why on earth would you want to use the polar form? One reason is that it's really easy to multiply and take powers in polar form. Imagine you wanted to multiply $3e^{i\pi/4}$ by $2e^{-i(3\pi/8)}$. This is pretty simple-you just use the normal exponential rules (see Section 9.1.1 in Chapter 9) to write

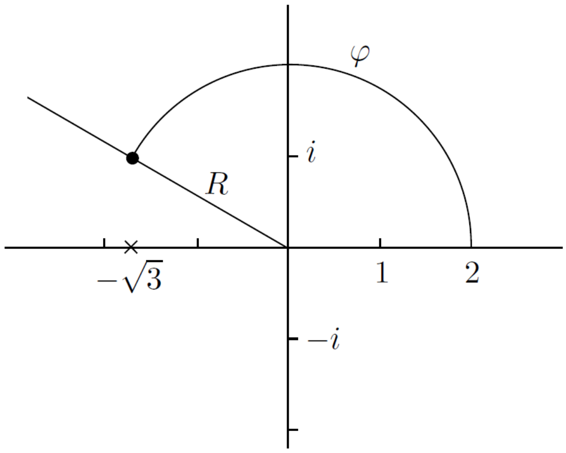
$$\left(3e^{i\pi/4}\right)\left(2e^{-i(3\pi/8)}\right)=6e^{-i(\pi/4-3\pi/8)}=6e^{-i\pi/8}$$

A lot of the time, you might want the final answer in Cartesian form. For example, suppose we'd like to compute $(1-i)^{99}$ and give the answer in Cartesian form. Expanding the expression by multiplying out would be crazy, so we won't go there. The correct way to proceed is to translate $1-i$ into polar form, take the 99th power, then translate back into Cartesian form (Page 603). In summary, to take a large power of a complex number, first convert it to polar form, then take the power. Find the largest even multiple of π less than the angle θ, and take that away from θ and replace θ by that new number. Finally, convert back to Cartesian form.

28.4 Solving $z^n = w$

Let's move onto a trickier subject: how to solve equations of the form $z^n = w$, where n is a given integer and w is a given complex number. This amounts to taking nth roots of w, but we don't just want to say $z = \sqrt[n]{w}$ since that doesn't tell us very much. Instead, we'll try to find a solution directly. Since powers work so well in polar form, that's what we'll use.

For example, to solve $z^5 = -\sqrt{3} + i$, we should use polar coordinates for both z and $w = -\sqrt{3} + i$. Since we don't know what z is, let's put $z = re^{i\theta}$. Now to find z, we just have to find what r and θ are. As for w, let's write $-\sqrt{3}+i=Re^{i\varphi}$ and then find R and φ. Now, let's draw a picture of the situation:



So our equation becomes

$$r^5 e^{i(5\theta)} = 2e^{i(5\pi/6)}$$

We can dissect the above equation into

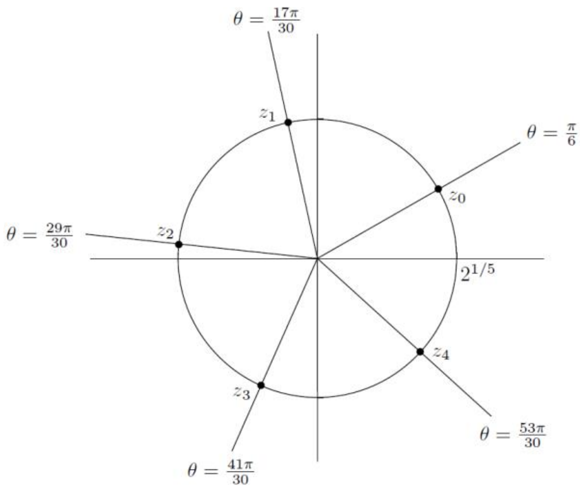
$$r^5 = 2 \quad \text{and} \quad e^{i(5\theta)} = e^{i(5\pi/6)}$$

The first is easy to solve: just take the 5th root to get $r = 2^{1/5}$, which is legit since r is a nonnegative real number. As for the second equation, you may be tempted to say $5\theta = 5\pi/6$, but it's not that simple. Remember, $e^{i\theta}$ is 2π- periodic in the variable θ! You can express this fact via the following important principle

if $e^{iA} = e^{iB}$ for real numbers A and B; then
$A = B + 2\pi k$, where k is an integer

So it looks as if there are infinitely many values of θ, and therefore infinitely many values of z that solve our equation. Appearances can be deceptive, however! You see, since $n = 5$, you only need to use the first five values for k, namely, $k = 0, 1, 2, 4$ (Page 605).

It's interesting to plot the solutions in the complex plane (Page 606)



In general, there are n solutions to the equation $z^n = w$ $\left(z = \sqrt[n]{w}\right)$, which when plotted form the vertices of a regular n-sided polygon. (The exception is if $w = 0$, in which case $z = 0$ is the only solution, but it is of multiplicity n.)

So, let's outline the main steps in solving $z^n = w$: (Page 607)

● Some variations

Suppose you want to solve the equation $(z-2)^3 = i$. No problem-just let $Z = z-2$, so that the equation is $Z^3 = i$. Solve this exactly as we just did at the end of the previous section.

Here's a tougher one (Page 609)

$$z^2 + \frac{1}{\sqrt{2}}z - \frac{\sqrt{3}i}{8} = 0$$

$$z = \frac{\frac{-1}{\sqrt{2}} \pm \sqrt{\frac{1}{2} + i\frac{\sqrt{3}}{2}}}{2}$$

We need to find the square roots of the complex number $\frac{1}{2} + i\frac{\sqrt{3}}{2}$. How do we do this? By solving the equation $Z^2 = \frac{1}{2} + i\frac{\sqrt{3}}{2}$.

One more example. How would you factor $(z^4 - z^2 + 1)$ over the complex numbers? Let's set $Z = z^2$, so that the equation becomes $Z^2 - Z + 1 = 0$ (Page 609)

$$z^4 - z^2 + 1 = \left(z - \frac{\sqrt{3}+i}{2}\right)\left(z - \frac{\sqrt{3}-i}{2}\right)\left(z - \frac{-\sqrt{3}+i}{2}\right)\left(z - \frac{-\sqrt{3}-i}{2}\right)$$

This is the complex factorization. To get the real factorization, we need to use a nice fact: if w is any complex number, then $(z-w)(z-\overline{w})$ has real coefficients when you multiply it out. Indeed, you get $z^2 - (w+\overline{w})z + w\overline{w}$, but it's easy enough to see that $w + \overline{w} = 2\,\mathrm{Re}(w)$ (which is real), and we've already seen that $w\overline{w} = |w|^2$, which is also real.

28.5 Solving $e^z = w$

Now it's time to see how to solve equations of the form $e^z = w$ for given w. It'd be nice if we could just write $z = \ln(w)$, but this isn't very helpful. For example, what exactly is $\ln(-\sqrt{3}+i)$? Let's try to answer this question.

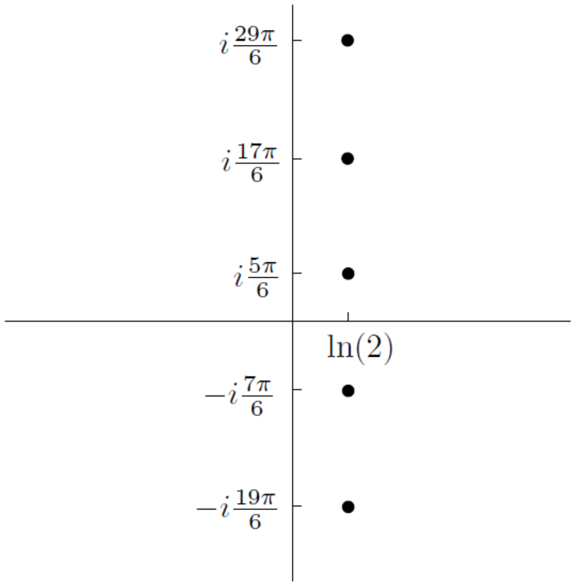
Fortunately, solving $e^z = w$ isn't much harder than solving $z^n = w$; in fact, if anything, it's simpler. Before we see how to do this, we need to understand e^z a little better. Let's see what happens if we write $z = x + iy$. We get

$$e^z = e^{x+iy} = e^x e^{iy}$$

So what? Well, the main point is that this is already in polar form. The modulus is e^x and the argument is y. If you prefer, $r = e^x$ (remember, e^x is real and positive) and $\theta = y$. This means that if z is in Cartesian form $x + iy$, then e^z is **automatically** in polar form: $e^z = e^x e^{iy}$.

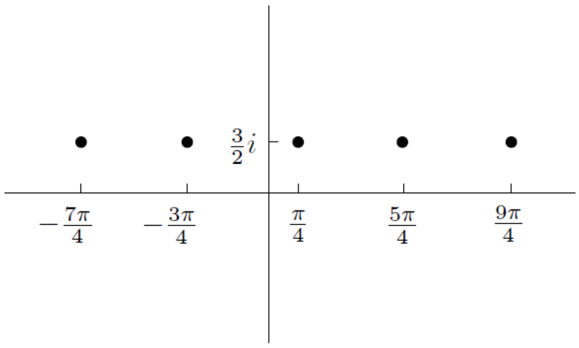
So, the main difference between solving $e^z = w$ and $z^n = w$ is that you don't need to put z in polar form in the first case, whereas you do in the second case. A sort of by-product of this is that there are infinitely many solutions to the equation $e^z = w$ (unless $w = 0$, in which case there are no solutions).

Let's solve $e^z = -\sqrt{3} + i$



The solutions are equally spaced on the vertical line. Incidentally, this means that they form an arithmetic progression of complex numbers.

Let's look at one more example. Suppose you want to solve $e^{2iz+3} = i$



Once again, the solutions are in arithmetic progression, but this time they lie on the horizontal line $y = \frac{3}{2}$.

28.6 Some Trigonometric Series

A trigonometric series is a series of the form

$$\sum_{n=0}^{\infty}(a_n \cos(n\theta) + b_n \sin(n\theta))$$

for some coefficients $\{a_n\}$ and $\{b_n\}$.

For example, consider the trigonometric series

$$\sum_{n=0}^{\infty}\frac{\sin(n\theta)}{n!}$$

where θ is real. Note that this is not a power series in θ, since $\sin(n\theta)$ is not a power of θ. On the other hand, we can make the whole thing into a power series by using the complementary series

$$\sum_{n=0}^{\infty}\frac{\cos(n\theta)}{n!}$$

in a clever way. In fact, we can find both series at once. The key is Euler's identity. Let's find both series at once by combining them like this:

$$\sum_{n=0}^{\infty}\frac{\cos(n\theta)}{n!}+i\sum_{n=0}^{\infty}\frac{\sin(n\theta)}{n!}$$

$$\sum_{n=0}^{\infty}\frac{\cos(n\theta)+i\sin(n\theta)}{n!}=\sum_{n=0}^{\infty}\frac{e^{in\theta}}{n!}=\sum_{n=0}^{\infty}\frac{\left(e^{i\theta}\right)^{n}}{n!}$$

Now, the last sum looks familiar. In fact, we saw in Section 28.1.1 above that

$$\sum_{n=0}^{\infty}\frac{z^n}{n!}=e^z$$

for all complex numbers z. Now we can get

$$\sum_{n=0}^{\infty}\frac{\left(e^{i\theta}\right)^{n}}{n!}=e^{e^{i\theta}}$$

So

$$\sum_{n=0}^{\infty}\frac{\cos(n\theta)}{n!}+i\sum_{n=0}^{\infty}\frac{\sin(n\theta)}{n!}=e^{e^{i\theta}}$$

Well, we need to convert the right-hand side into Cartesian form

$$e^{e^{i\theta}}=e^{\cos(\theta)+i\sin(\theta)}=e^{\cos(\theta)}e^{i\sin(\theta)}$$

This is the polar form of $e^{e^{i\theta}}$. To get the Cartesian form, we need to convert $e^{i\sin(\theta)}$ into $\cos(\sin(\theta))+i\sin(\sin(\theta))$. Putting it all together, we get

$$\sum_{n=0}^{\infty}\frac{\cos(n\theta)}{n!}+i\sum_{n=0}^{\infty}\frac{\sin(n\theta)}{n!}=e^{\cos(\theta)}\cos(\sin(\theta))+ie^{\cos(\theta)}\sin(\sin(\theta))$$

Now, if two complex numbers are equal, then their real parts must be equal, and also their imaginary parts must be equal

$$\sum_{n=0}^{\infty}\frac{\cos(n\theta)}{n!}=e^{\cos(\theta)}\cos(\sin(\theta))\quad\text{and}\quad\sum_{n=0}^{\infty}\frac{\sin(n\theta)}{n!}=e^{\cos(\theta)}\sin(\sin(\theta))$$

Not easy, but this is basically what you have to do. I'll do one more example (Page 614), to find (we will have a geometric series)

$$\sum_{n=0}^{\infty}\frac{\cos(n\theta)}{3^n}\quad\text{and}\quad\sum_{n=0}^{\infty}\frac{\sin(n\theta)}{3^n}$$

28.7 Euler's Identity and Power Series

Let's finish the chapter with a justification of Euler's identity

$$e^{i\theta}=\cos(\theta)+i\sin(\theta)$$

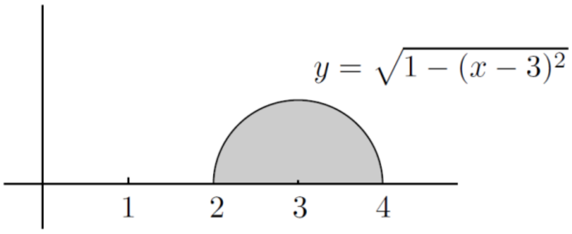
using power series (Page 615).

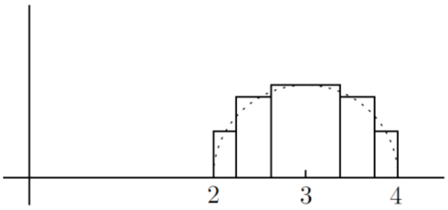
CHAPTER 29 Volumes, Arc Lengths, and Surface Areas

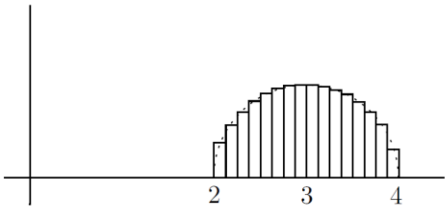
We have used definite integrals to find areas. Now we're going to use them to find volumes, lengths of curves, and surface areas.

29.1 Volumes of Solids of Revolution

We'll start with finding volumes of solids of revolution. The idea is that there is some region in the plane, and some axis also in the plane, and a solid is formed by revolving the region about the axis. For our purposes, the axis will always be parallel to the x-axis or the y-axis. (It is possible to deal with diagonal axes, but it's a real pain unless you use techniques from linear algebra.)



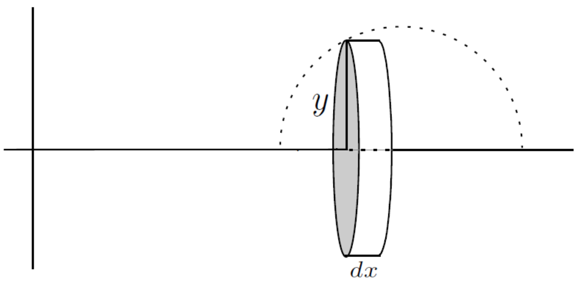




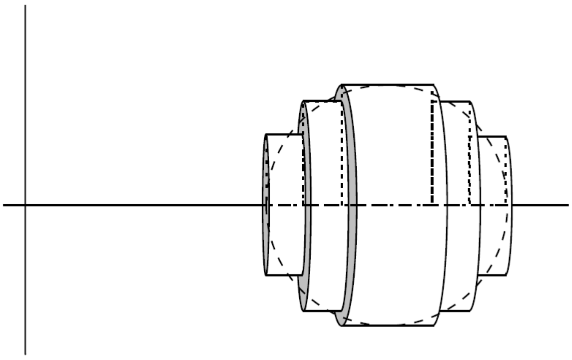
Here's the pattern: we make a little strip of width dx units and height y units at position x on the x-axis, work out its area, then put a definite integral sign in front to get the total area we're looking for. This technique doesn't just work for areas-it also works for volumes. In particular, let's see how it works using two different methods for finding volumes of revolution: the disc method and the shell method.

● The disc method

Suppose that we revolve the semicircle from the previous section about the x-axis. This will give us a sphere. Let's try to work out its volume. We'll start with one strip, just like in the picture at the end of the previous section, and revolve that strip about the x-axis. Here's what we get:



This is a thin disc of width dx units and radius y units. Since the volume of a cylinder of radius r units and height h units is $\pi r^2 h$ cubic units, the volume of our thin disc is $\pi y^2 dx$ cubic units. For example, if you use five strips, you might get something like this:



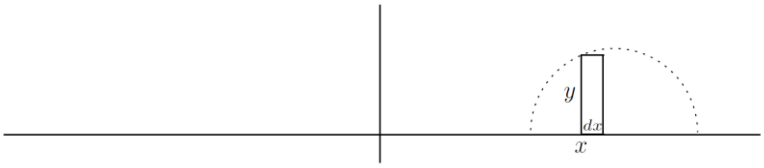
In the limit, as the maximum disc thickness goes down to zero, the approximation becomes perfect: the total volume of the discs tends toward the volume of the sphere. In our case, $y = \sqrt{1-(x-3)^2}$ and x goes from 2 to 4, so we have

$$V = \int_2^4 \pi y^2 dx \ = \pi \int_2^4 (1-(x-3)^2) dx$$

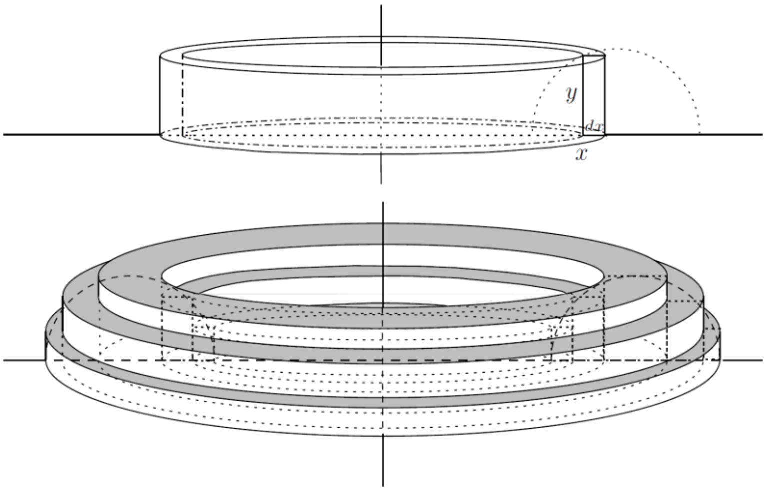
The volume works out to be $\frac{4}{3}\pi$ cubic units (try it!) which is what we'd expect, since we're dealing with a sphere of radius 1 unit here. The method we just used is called the *disc method*; it is also known as the method of *slicing*.

● The shell method

Now, let's suppose that we take our favorite semicircular region from before but this time we revolve it about the y-axis. Try to imagine what you'd get. Let's approximate the semicircle by thin strips again, but this time we'll revolve each strip about the y-axis, instead of the x-axis. As we saw before, a typical strip looks like this:

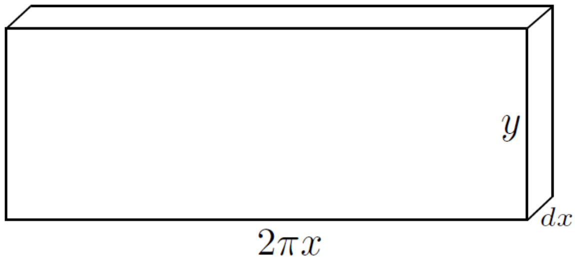


When you revolve it about the y-axis, you don't get a disc-you get a cylindrical shell:



This weird solid is a pretty lumpy bagel half, but its volume is fairly close to what we're looking for

First we need to find the volume of one generic shell. The point is, it's **almost** a rectangular prism. So, the idealized version of the unfolded can looks like this:



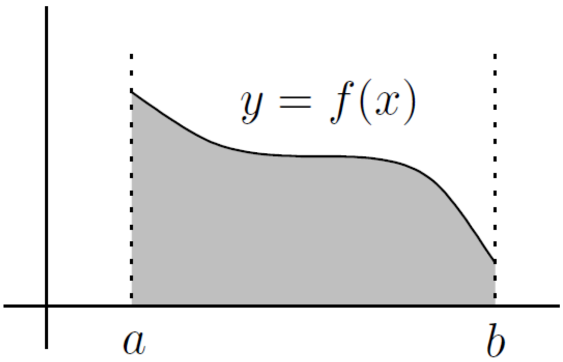
The thickness is dx units, and the side we cut along is still the height of the cylindrical shell, that is, y units. The long side is equal to the circumference of the shell (think about it!) which is $2\pi x$ units, since the radius of the shell is basically x units. The volume of the shell is very close to $2\pi xydx$ cubic units. Now all we have to do is integrate from $x = a$ to $x = 4$ to see that the volume of the bagel half (in cubic units) is (Page 621)

$$\int_2^4 2\pi xydx = 2\pi \int_2^4 x\sqrt{1-(x-3)^2}dx$$

The method we just used is, unsurprisingly, called the *shell method*.

● Summary . . . and variations

So far we have seen how to use the disc and shell methods in the special case of our semicircle. The same method works for general regions which are contained between a curve, the x-axis, and two vertical lines:



By the same reasoning that we used above in the special case of the semicircle, we can arrive at the following principles:

● If you revolve the area under the curve $y = f(x)$ between $x = a$ and $x = b$ (as shown above) about the x-axis, then the disc method applies and the volume is equal to

$$\int_a^b \pi y^2 dx \quad \text{cubic units}$$

● If you revolve the area under the curve $y = f(x)$ between $x = a$ and $x = b$ (as shown above) about the y-axis, then the shell method applies and the volume is equal to

$$\int_a^b 2\pi xydx \quad \text{cubic units}$$

It's an even better idea to be able to derive these formulas by knowing how to find the volume of a typical disc or shell. This will be especially useful if you encounter one (or more) of the following variations:

1. The region to be revolved might lie between a curve and the y-axis (instead of the x-axis)

2. The region to be revolved might lie between two curves, instead of just being a region under a curve down to an axis

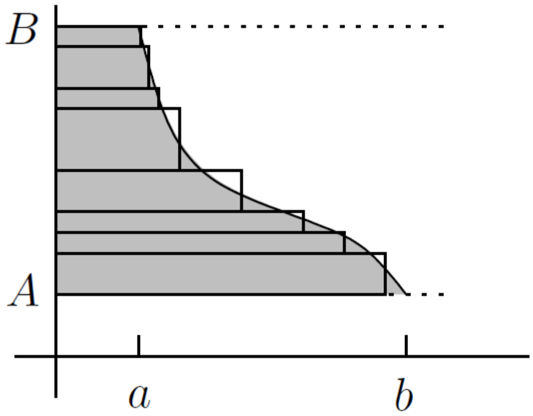
3. The axis of revolution may be parallel to the x-axis or y-axis, not the axis itself

It's important to know how to decide whether to use the disc method or the shell method:

- if the really thin bit of each strip is **parallel** to the axis of revolution, the **disc method** applies
- whereas if the really thin bit of each strip is **perpendicular** to the axis of revolution, the **shell method** applies

● Variation 1: regions between a curve and the y-axis

If the region is between the curve and the y-axis, you probably want to take strips lying on their sides, with the thin part of the strip along the y-axis:

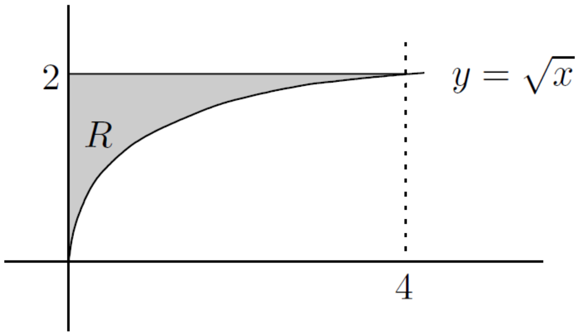


Be very careful that the limits of integration are relevant points on the y-axis, not the x-axis, since the integral is taken with respect to y (because of the dy).

In summary, then, the rule of thumb is this:

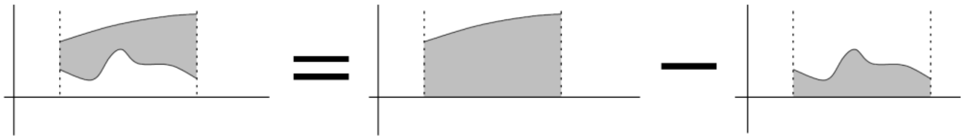
If the region lies between a curve and the y-axis, switch x and y

Here's an example of Variation 1 (Page 624)

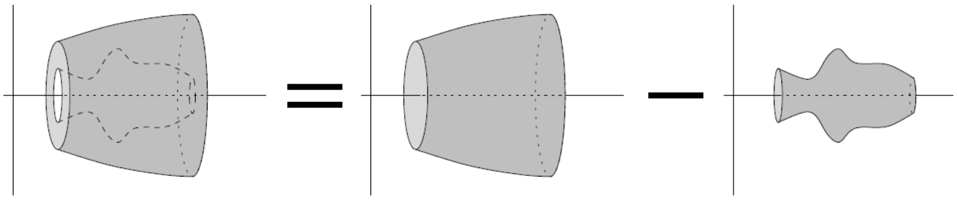


● Variation 2: regions between two curves

Suppose the region to be revolved lies between two curves. We'll handle this situation in the same way as finding the area of a region between two curves in Section 16.4.2 of Chapter 16. The general idea is to take the top curve and revolve the region under it all the way to the axis, to get a bigger solid than you want. Now take the bottom curve and revolve the region under it all the way to the axis, to get a solid which you actually need to cut out of the big solid and throw away to get the desired solid. Finally, subtract the small volume from the big one. Indeed, consider the following three regions:



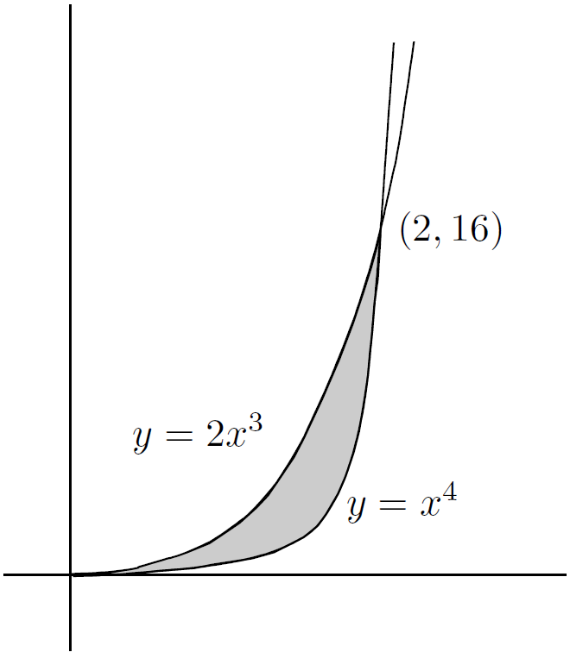
Now, regardless of whether you revolve about the x-axis or the y-axis, the volume of revolution of the region we want is equal to the difference between the volume of revolution of the big region and the volume of revolution of the small region



So, here's what we conclude:

If the region lies between two curves, find the difference between the two corresponding volumes of revolution

Let's look at a concrete example. What is the volume of the solid formed by revolving the region about the x-axis?

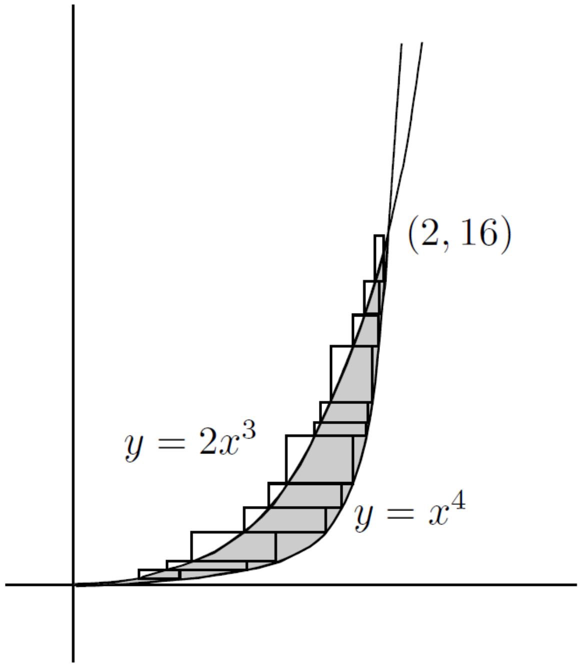


Use the disc method on each of the two curves to see that the volume we want is

$$\int_0^2 \pi y_1^2 dx - \int_0^2 \pi y_2^2 dx = \pi \int_0^2 (2x^3)^2 dx - \pi \int_0^2 (x^4)^2 dx$$

You should work this out and check that the answer is $1024\pi/63$ cubic units.

How about revolving the same region about the y-axis? First, the disc method:

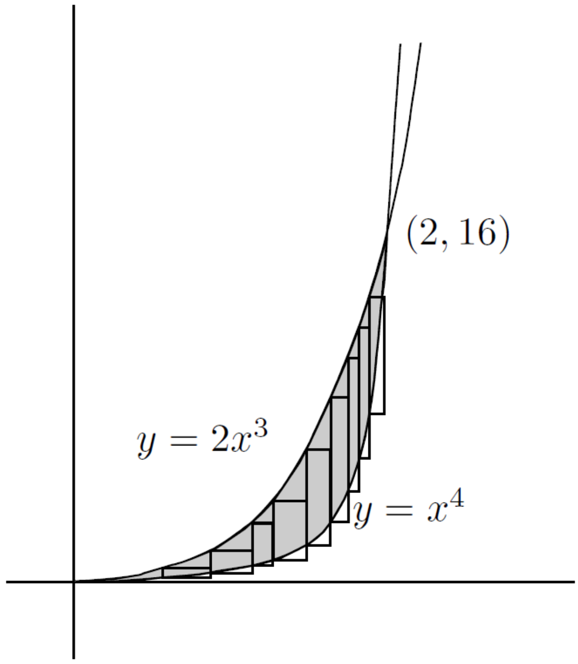


$$\int_0^{16} \pi x_1^2 dy - \int_0^{16} \pi x_2^2 dy = \pi \int_0^{16} \left(y^{1/4}\right)^2 dy - \pi \int_0^{16} \left((y/2)^{1/3}\right)^2 dy$$

$$= \pi \int_0^{16} y^{1/2} dy - 2^{-2/3}\pi \int_0^{16} y^{2/3} dy$$

This works out to be $64\pi/15$.

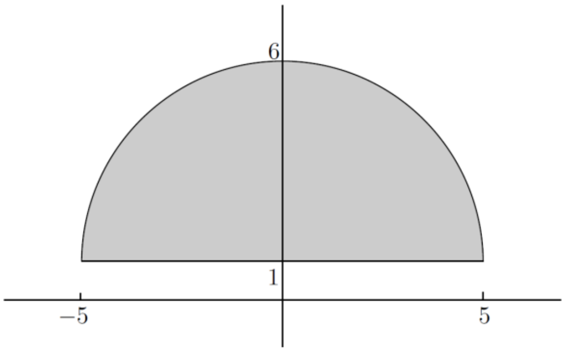
Let's try to find the same volume by using shells. This time, we slice the region vertically:



$$\int_0^2 2\pi x y_1 dx - \int_0^2 2\pi x y_2 dx = 2\pi \int_0^2 2x^4 dx - 2\pi \int_0^2 x^5 dx$$

which is $64\pi/15$ cubic units.

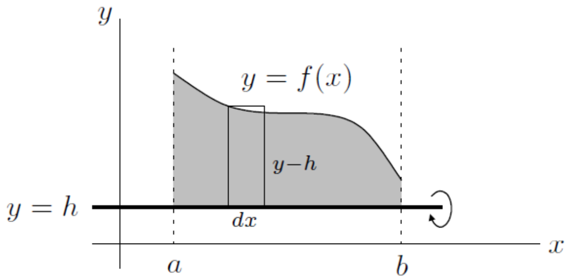
This variation also applies when the area doesn't go all the way down to the axis. For example, suppose we want to find the volume of revolution when the region between the curve $y = 1 + \sqrt{25 - x^2}$ and the line $y = 1$ is revolved about the x-axis



$$\int_{-5}^{5} \pi \left(1 + \sqrt{25 - x^2}\right)^2 dx - \int_{-5}^{5} \pi (1)^2 dx = 25\pi^2 + 500\pi/3$$

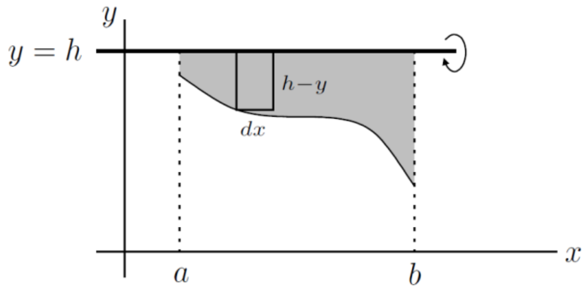
● Variation 3: revolving about axes parallel to the coordinate axes

Finally, let's see how to handle revolution about the axis $x = h$ or $y = h$, where h is some number not necessarily equal to 0. We'll start with $y = h$, which is parallel to the x-axis but is at height h. Suppose we want to revolve the region between the curve $y = f(x)$ and the lines $y = h$, $x = a$, and $x = b$ about the line $y = h$:

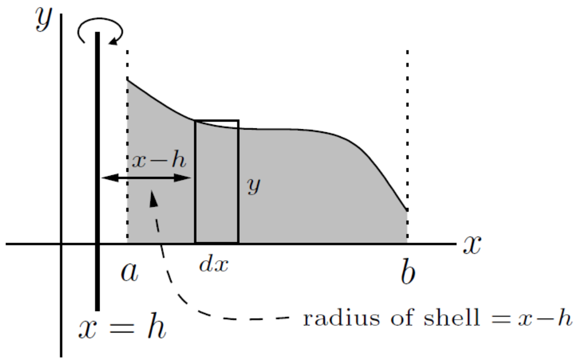


The width is dx, but the height isn't y: it's $y - h$. If h happens to be negative, then the height of the strip is more than y. . . The volume of the whole solid of revolution is $\int_a^b \pi(y - h)^2 dx$.

In fact, the only difference between this formula and the regular disc method is that y has been replaced by the quantity $(y - h)$. The only problem with this is that it's possible that the line $y = h$ is actually above the curve, like this:

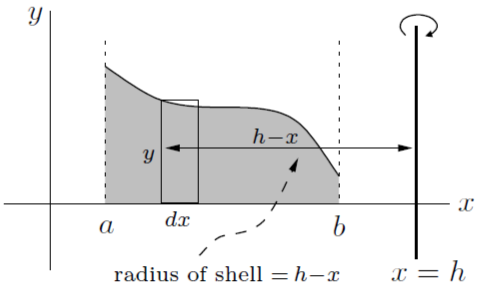


In this case, the height of the strip is $h - y$, not $y - h$. It's good to be careful about these things. Besides, the shell method is a different story. Suppose we now want to find the volume of the solid formed by revolving the region below about the axis $x = h$:



The total volume is $\int_a^b 2\pi(x-h)y\,dx$ cubic units.

How about if the axis is to the right of the region? Consider the following picture:

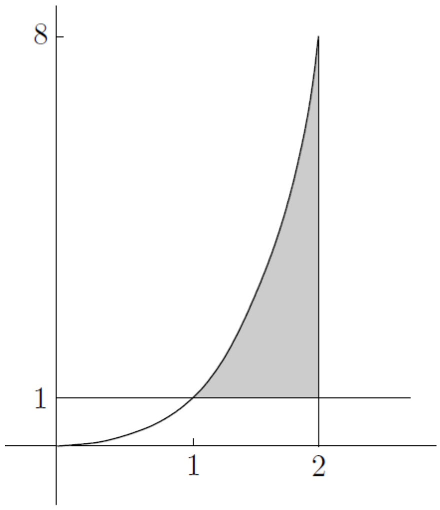


The volume of revolution is $\int_a^b 2\pi(h-x)y\,dx$ cubic units.

So, here's the general idea for Variation 3:

If the axis of revolution is $x = h$, replace x by $(x-h)$ (or $(h-x)$ if $x < h$)
If the axis of revolution is $y = h$, replace y by $(y-h)$ (or $(h-y)$ if $y < h$)

Let's look at some examples of Variation 3. In all the examples, we'll be dealing with the region between the curve $y = x^3$, the line $x = 2$ and the line $y = 1$:



Let's start with finding the volume when the region is revolved about the line $y = 1$. The volume is given by

$$\int_1^2 \pi(y-1)^2 dx = \pi\int_1^2 (x^3-1)^2 dx$$

which easily works out to be $163\pi/14$ cubic units.

How about revolving the same region about the line $x=2$? This is actually a combination of Variation 1 and Variation 3. The volume is

$$\int_1^8 \pi(2-x)^2 dy = \pi \int_1^8 \left(2-y^{1/3}\right)^2 dy$$

which simplifies to $8\pi/5$ cubic units. It's a good idea to make sure that you can also work this out by finding the volume of a typical disc.

Now, what about if we revolve the same region about $x=-3$? We'll use a combination of Variation 2 and Variation 3. The volume is given by

$$\int_1^2 2\pi(x+3)y_1 dx - \int_1^2 2\pi(x+3)y_2 dx = 2\pi \int_1^2 (x+3)x^3 dx - 2\pi \int_1^2 (x+3) dx$$

which works out to be $259\pi/10$ cubic units.

Let's repeat the same example, this time taking horizontal strips. Now we have to use the disc method. The volume is

$$\int_1^8 \pi(x_1+3)^2 dy - \int_1^8 \pi(x_2+3)^2 dy = \pi \int_1^8 (2+3)^2 dy - \pi \int_1^8 \left(y^{1/3}+3\right)^2 dy$$

which again works out to be $259\pi/10$ cubic units.

29.2 Volumes of General Solids

Most solids can't be formed by revolving some planar area about an axis in that plane. One technique for finding the volume of such a solid is the method of *slicing*, which actually generalizes the disc method from Section 29.1.1 above

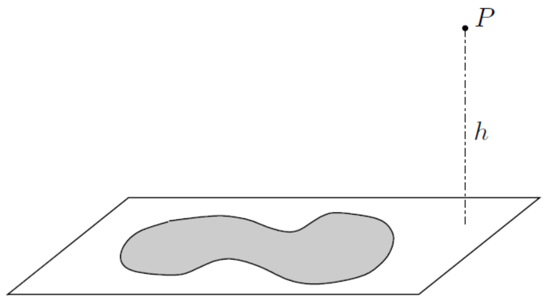
Basically, here is your choice: (Page 632)

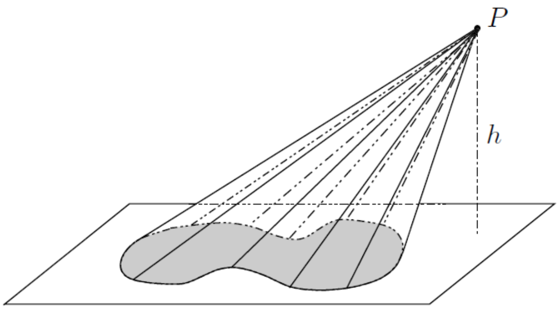
1. Choose an axis
2. Find a typical cross-sectional area at a point x on the axis; call this area $A(x)$ square units
3. Then if V is the volume of the solid (in cubic units), we have

$$V = \int_a^b A(x) dx$$

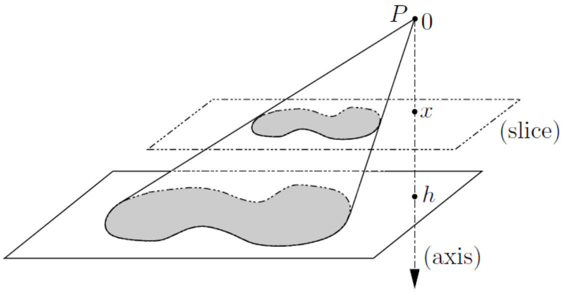
where $[a,b]$ is the range of x which completely covers the solid.

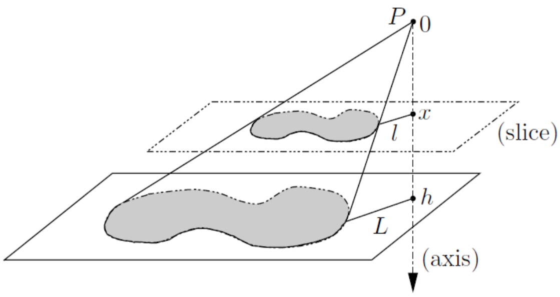
Let's use the above technique to find the volume of a "generalized" cone (Page 632)

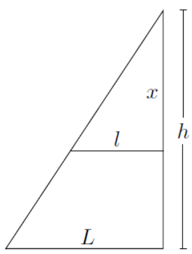




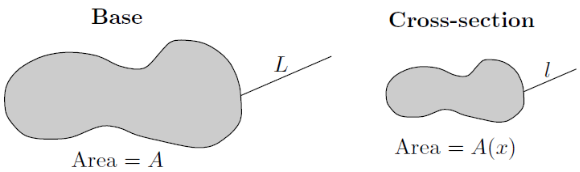
So, how do we find the volume?







Using similar triangles, we can get $l = xL/h$



Now here's an important principle of similarity: the ratio of areas of the figures is the **square** of the ratio of the two corresponding lengths

$$\frac{A}{A(x)} = \left(\frac{L}{l}\right)^2$$

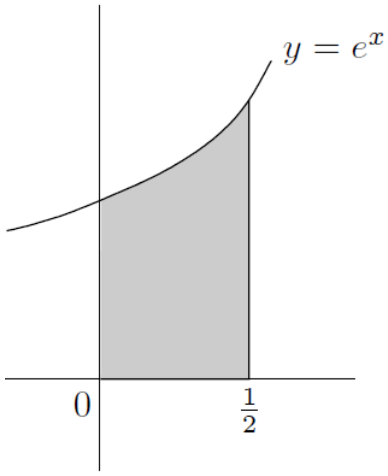
$$A(x) = \frac{Al^2}{L^2} = \frac{A}{L^2} \cdot \left(\frac{xL}{h}\right)^2 = \frac{Ax^2}{h^2}$$

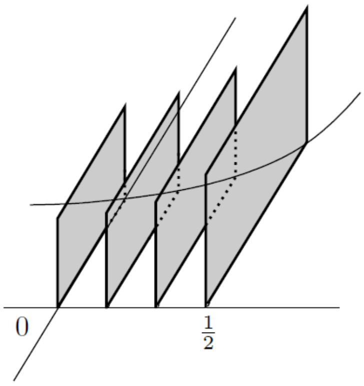
Finally, we're ready to integrate!

$$V = \int_0^h A(x)dx = \int_0^h \frac{Ax^2}{h^2} dx = \frac{A}{h^2}\int_0^h x^2 dx = \frac{A}{h^2} \cdot \frac{h^3}{3} = \frac{1}{3}Ah \text{ (cubic units)}$$

Hey, so we just got the formula for the volume of any sort of pyramid or cone-like object. For example, for the regular old cone, the volume is $\frac{1}{3}\pi r^2 h$ cubic units, which is exactly what we found above since $A = \pi r^2$. Same thing for a square pyramid-the volume is $\frac{1}{3}l^2 h$ cubic units (where the side length of the base is l units), which works as well because the base area is given by $A = l^2$.

Let's look at one more example (Page 636)

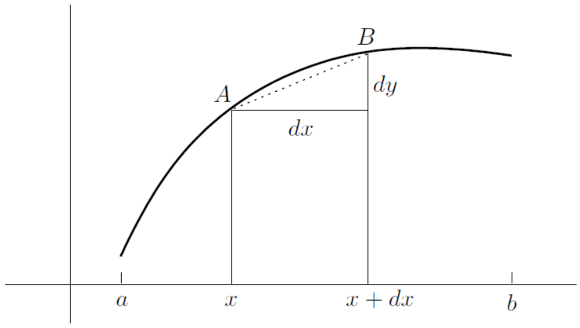




29.3 Arc Lengths

Say we have a graph of $y = f(x)$ for some function f, where x ranges from a to b. This length is called the *arc length* of the curve, and we're going to find a formula for it. The strategy will be to get a sort of prototype expression, then to adapt this to get several useful versions of the formula.

So, let's look at a little piece of curve between x and $x + dx$:



Let's approximate the length of the curve between A and B by the length of the dotted line segment AB. The closer A and B are to each other, the better the approximation. By Pythagoras' Theorem, the length of AB is $\sqrt{(dx)^2 + (dy)^2}$ units. Now we just need to repeat this process with lots of little line segments. As usual, the integration takes care of the adding up and limiting parts, but you have to be careful. If you just put an integral sign in front of the little length $\sqrt{(dx)^2+(dy)^2}$, you'll get

$$\text{arc length} = \int_?^? \sqrt{(dx)^2+(dy)^2}$$

The problem is, this integral doesn't really mean anything!
We need to integrate with respect to one variable. Anyway, in each of the cases below, we'll see how to adapt the above prototypical formula to get a legitimate formula for arc length:
1. If $y=f(x)$ and x ranges from a to b, then take out a factor of $(dx)^2$ in the above integrand (as we just did above) and pull it out of the square root to get

$$\boxed{\text{arc length} = \int_a^b \sqrt{1+\left(\frac{dy}{dx}\right)^2}\,dx} \qquad \text{(standard form)}$$

In terms of f, you can rewrite this as

$$\text{arc length} = \int_a^b \sqrt{1+\left(f'(x)\right)^2}\,dx$$

2. Suppose that x is given in terms of y. If $x=g(y)$ and y ranges from A to B, then you take out a factor of $(dy)^2$ instead (or if you prefer, swap each occurrence of x and y in the boxed formula above) to get

$$\boxed{\text{arc length} = \int_A^B \sqrt{1+\left(\frac{dx}{dy}\right)^2}\,dy} \qquad \text{(in terms of } y\text{)}$$

which can also be written as

$$\text{arc length} = \int_A^B \sqrt{1+\left(g'(y)\right)^2}\,dy$$

3. How about the parametric form? This means that x and y are functions of a parameter t which ranges from t_0 to t_1. We can think of the quantity $(dx)^2$ as $(dx/dt)^2(dt)^2$ and similarly for y. We can then pull the $(dt)^2$ out and take its square root to get the useful formula:

$$\boxed{\text{arc length} = \int_{t_0}^{t_1} \sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2}\,dt} \qquad \text{(parametric version)}$$

4. A special case of this last formula occurs in the case of polar coordinates. The curve is $r=f(\theta)$, where θ ranges from θ_0 to θ_1; We know that $x=r\cos(\theta)$ and $y=r\sin(\theta)$, so replacing r by $f(\theta)$. Finally

$$\boxed{\text{arc length} = \int_{\theta_0}^{\theta_1} \sqrt{\left(f(\theta)\right)^2+\left(f'(\theta)\right)^2}\,d\theta} \qquad \text{(polar, } r=f(\theta)\text{)}$$

By the way, you should express all these arc lengths in units.
Let's look at some examples (Page 639)

- Parametrization and speed

Before we move on to finding surface areas, there's one little fact related to the arc length formula in parametric coordinates that I'd like to look at. Suppose an ant (not a snail, this time!) is crawling around on a flat piece of ground, and we define the ant's position at time t seconds to be $(x(t), y(t))$. The ant's velocity in the x direction is dx/dt and its velocity in the y direction is dy/dt. Its real speed has to involve both of these velocities, by Pythagoras' Theorem, we should have:

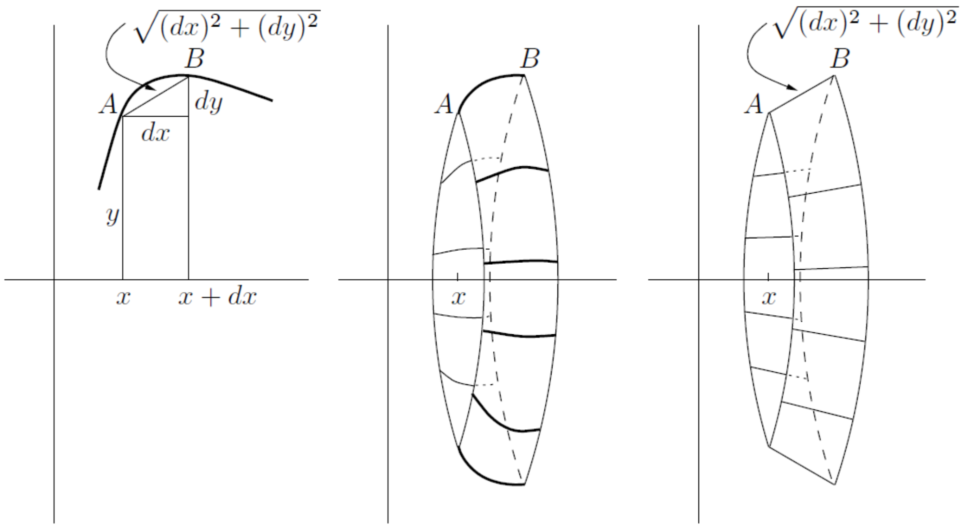
$$\boxed{\text{speed} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}}$$

Hey, this is the quantity that we've been integrating to find arc length in the parametric case! So we now have a meaning for the integrand in the formula for arc length, at least in the parametric case: it is the instantaneous speed of a particle moving along the curve.

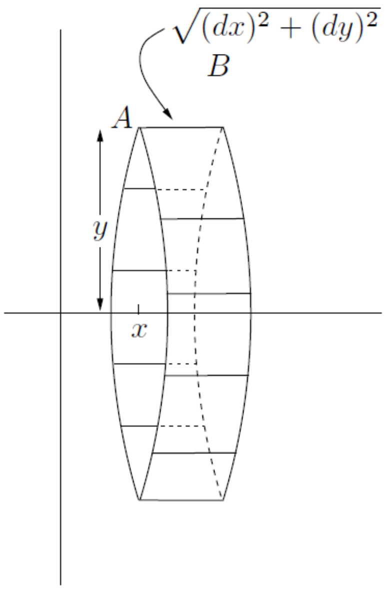
Consider some examples (Page 640)

29.4 Surface Areas of Solids of Revolution

The last thing we'll consider in this chapter is how to find the surface area of a surface formed by revolving a curve about an axis. The method is a sort of combination of how we found arc lengths and volumes. Let's suppose we are revolving about the x-axis (Page 641)



Actually, we are even lazier than that: now the loop is cylindrical



So we are led to the prototypical formula for revolution about the x-axis:

$$\text{surface area} = \int_{?}^{?} 2\pi y \sqrt{(dx)^2 + (dy)^2} \quad (\text{revolution about } x - \text{axis})$$

The prototypical formula for revolution about the y-axis is

$$\text{surface area} = \int_{?}^{?} 2\pi x \sqrt{(dx)^2 + (dy)^2} \quad (\text{revolution about } y - \text{axis})$$

Let's see how we can modify the formulas so we can actually use them:

1. Suppose we want to revolve the curve $y = f(x)$ about the x-axis, where x ranges from a to b. We take out a factor of $(dx)^2$ in the integrand of the first prototypical formula and pull it out of the square root, just as we did in the case of arc length, to get

$$\boxed{\text{surface area} = \int_{a}^{b} 2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx}$$

(about the x-axis)

In terms of f, it looks like this:

$$\text{surface area} = \int_{a}^{b} 2\pi f(x) \sqrt{1 + \left(f'(x)\right)^2}\, dx$$

2. If instead we want to revolve the same curve about the y-axis, the same manipulations applied to the other prototypical formula give

$$\boxed{\text{surface area} = \int_{a}^{b} 2\pi x \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx}$$

(about the y-axis)

or in terms of f,

$$\text{surface area} = \int_{a}^{b} 2\pi x \sqrt{1 + \left(f'(x)\right)^2}\, dx$$

3. Of course, there's also a parametric form. If x and y are functions of a parameter t which ranges from t_0 to t_1, then dividing and multiplying by dt leads to the following formulas:

$$\boxed{\text{surface area} = \int_{t_0}^{t_1} 2\pi y \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt}$$

(parametric version, about the x-axis)

and

$$\boxed{\text{surface area} = \int_{t_0}^{t_1} 2\pi x \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt}$$

(parametric version, about the y-axis)

Again, all of these surface areas are in square units.

Here're some examples (Page 643).

CHAPTER 30 Differential Equations

A differential equation is an equation involving derivatives. These things are really useful for describing how quantities change in the real world.

30.1 Introduction to Differential Equations

We considered the equation

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

where k is some fixed constant, and claimed that the only solutions to it are of the form $y = Ae^{kx}$ for some constant A. The equation $dy/dx = ky$ is an example of a ***first**-order differential equation*. This is because there's only a first derivative floating around. In general, the order of a differential equation is the order of the highest derivative involved. For example, a fourth-order differential equation (Page 646).

Now consider a specific example of the first-order differential equation at the beginning of this section, but with an extra condition:

$$\frac{dy}{dx} = -2y, \quad y(0) = 5$$

We know $y = Ae^{kx}$ is the general solution, so set $k = -2$. Now put $x = 0$ and $y = 5$ to see that $5 = Ae^{-2(0)}$, get $A = 5$. So the actual solution is $y = 5e^{-2x}$.

What we have just been looking at is an example of an *initial value problem*, or *IVP*. The idea is that you know a starting condition (in this case, $y(0) = 5$) as well as a differential equation that tells you how the situation evolves from there (in this case, $dy/dx = -2y$), and you can use these two facts to find out the exact solution with no undetermined constants. For a second-order differential equation, you effectively need to integrate twice, so you'll get two undetermined constants; it follows that you need **two** pieces of information.

Now, the study of differential equations is pretty bloody huge. These things are hard to solve. Luckily, there are some simple types which can be solved without too much trouble.

30.2 Separable First-order Differential Equations

A first-order differential equation is called ***separable*** if you can put all the y-stuff on one side (including the dy), and all the x-stuff on the other side (including the dx).

For example, the equation $dy/dx = ky$ can be rearranged to read (Then integrate both side)

$$\frac{1}{ky}dy = dx$$

so it is separable. As another example, the equation

$$\frac{dy}{dx} + \cos^2(y)\cos(x) = 0$$

can be rearranged (check out the algebra yourself!) into

$$\sec^2(y)\,dy = -\cos(x)\,dx$$

(Page 646)

30.3 First-order Linear Equations

Here's a different type of first-order differential equation:

$$\frac{dy}{dx} + p(x)y = q(x)$$

where p and q are given functions of x. Such an equation is called a ***first-order linear*** differential equation. It may not be separable, and it may not even look particularly linear! For example,

$$\frac{dy}{dx} + 6x^2 y = e^{-2x^3}\sin(x)$$

doesn't look very linear, yet this equation is indeed first-order linear. The reason is that the powers of y and dy/dx are both **one**.

Let's see the linear equation from above,

$$\frac{dy}{dx} + 6x^2 y = e^{-2x^3}\sin(x)$$

There's a neat trick: Imagine that we multiply both sides by the quantity e^{2x^3}

$$e^{2x^3}\frac{dy}{dx} + 6x^2 e^{2x^3} y = \sin(x)$$

Watch carefully, now: there's nothing up my sleeve as I rewrite this as

$$\frac{d}{dx}\left(e^{2x^3} y\right) = \sin(x)$$

Now all we have to do is integrate both sides with respect to x

$$e^{2x^3} y = \int \sin(x)\,dx = -\cos(x) + C$$

Dividing by e^{2x^3}, we get the solution

$$y = (C - \cos(x))e^{-2x^3}$$

where C is an arbitrary constant.

The key to the previous solution was multiplying by e^{2x^3}. The quantity e^{2x^3} is called *an integrating factor*. It turns out that for the general first-order linear differential equation

$$\frac{dy}{dx} + p(x)y = q(x)$$

a good integrating factor is given by the equation

$$\text{integrating factor} = e^{\int p(x)dx}$$

where you **don't** need a $+C$ in the integral. After you multiply the original differential equation by this integrating factor, the left-hand side can be "factored" as

$$\frac{d}{dx}(\text{integrating factor} \times y)$$

Here are two more examples. First, how would you solve

$$\frac{dy}{dx} = e^x y + e^{2x}, \quad y(0) = 2(e-1)?$$

(Page 650)

One more example of a first-order linear differential equation: (Page 651)

$$\tan(x)\frac{dy}{dx} = e^{\sin(x)} - y$$

In summary, here's the method for dealing with first-order linear differential equations:

- Put the stuff involving y on the left-hand side and the stuff involving x on the right-hand side, then divide through by the coefficient of dy/dx to get the equation into the standard form

$$\frac{dy}{dx} + p(x)y = q(x)$$

- Multiply through by the integrating factor, which we'll call $f(x)$, given by

$$\boxed{\text{integrating factor } f(x) = e^{\int p(x)dx}}$$

where no $+C$ is needed in the integral in the exponent

- The left-hand side becomes $\frac{d}{dx}(f(x)y)$, where $f(x)$ is the integrating factor. Rewrite the equation with this new left-hand side
- Integrate both sides; this time you must put a $+C$ on the right-hand side
- Divide by the integrating factor to solve for y

● Why the integrating factor works

Why is the weird expression $e^{\int p(x)dx}$ a good integrating factor? (Page 652)

30.4 Constant-coefficient Differential Equations

Now it's time to look at linear differential equations with **constant** coefficients. These equations look something like this:

$$a_n\frac{d^ny}{dx^n}+\cdots+a_2\frac{d^2y}{dx^2}+a_1\frac{dy}{dx}+a_0y=f(x)$$

Here f is some function of x only, and $a_n,\cdots,a_1,a_0$ are just plain old constant real numbers. Notice that the left-hand side of the above equation looks a bit like a polynomial in y, except that instead of taking powers of y, we are taking derivatives.

First, we need to look at some general ideas for solving both first- and second-order constant-coefficient linear equations.

Let's start by considering a simple case: assume there's no stuff in x on the right-hand side. Two such examples are

$$\frac{dy}{dx}-3y=0 \quad \text{and} \quad \frac{d^2y}{dx^2}-\frac{dy}{dx}+20y=0$$

Such equations are called *homogeneous*.

● Solving first-order homogeneous equations

This is pretty easy. The solution to

$$\frac{dy}{dx}+ay=0$$

is just $y=Ae^{-ax}$.

● Solving second-order homogeneous equations

This case is a little more involved. We need to solve

$$a\frac{d^2y}{dx^2}+b\frac{dy}{dx}+cy=0$$

Although it might seem a little strange, the easiest way to do this is to pluck a quadratic equation seemingly out of thin air. The quadratic equation, called the ***characteristic quadratic equation***, is $at^2+bt+c=0$. The next thing is to find the roots of the characteristic quadratic. There are three possibilities, depending on whether there are two real roots, one (double) real root or two complex roots. Let's summarize the whole method, then solve the above three examples.

How to solve the homogeneous equation $\boldsymbol{ay''+by'+cy=0}$:

1. Write down the characteristic quadratic equation $at^2+bt+c=0$ and solve it for t
2. If there are two different real roots α and β, the solution is

$$y=Ae^{\alpha x}+Be^{\beta x}$$

3. If there is only one (double) real root α, the solution is

$$y = Ae^{\alpha x} + Bxe^{\alpha x}$$

4. If there are two complex roots, they will be conjugate to each other. That is, they must be of the form $\alpha \pm i\beta$. The solution is

$$y = e^{\alpha x}(A\cos(\beta x) + B\sin(\beta x))$$

In all three cases (2, 3 and 4), A and B are undetermined constants

So, for example $y'' - y' - 20y = 0$, we saw that the characteristic quadratic equation is $t^2 - t - 20 = 0$. Factor the quadratic as $(t+4)(t-5)$, we see that the solution to our equation is given by

$$y = Ae^{-4x} + Be^{5x}$$

for some constants A and B

The characteristic quadratic equation $t^2 + 6t + 9 = 0$ in example $y'' + 6y' + 9y = 0$ reduces to $(t+3)^2 = 0$, so the solution to the homogeneous equation is

$$y = Ae^{-3x} + Bxe^{-3x}$$

Finally, if we use the quadratic formula to solve the characteristic quadratic equation $t^2 - 2t + 5 = 0$ of example $y'' - 2y' + 5y = 0$, we get $t = 1 \pm 2i$. So, with $\alpha = 1$ and $\beta = 2$, that the solution to the equation is

$$y = e^x(A\cos(2x) + B\sin(2x))$$

● Why the characteristic quadratic method works

Step 2 (Page 655)

Step 4 (Page 656)

Step 3 (Page 656)

● Nonhomogeneous equations and particular solutions

Now let's see what happens if we do have some stuff in x alone, which we put on the right-hand side. For example, consider the differential equation

$$y'' - y' - 20y = e^x$$

We know that the derivatives of e^x are all e^x, so let's try $y = e^x$. We will get $e^x - e^x - 20e^x = e^x$, which is not equal to the right-hand side, but it's pretty close. So, let's try again: set $y = -\frac{1}{20}e^x$, so we have

$$y'' - y' - 20y = -\frac{1}{20}e^x - \left(-\frac{1}{20}e^x\right) - 20\left(-\frac{1}{20}e^x\right) = e^x$$

So we have shown that $y = -\frac{1}{20}e^x$ is a solution to our original equation. It's not the only solution, though. For the related homogeneous equation $y'' - y' - 20y = 0$, we'll write $y = Ae^{-4x} + Be^{5x}$ as y_H instead of just y, where the H stands for homogeneous.

Here I wrote the solution $-\frac{1}{20}e^x$ from above as y_P; this is called a ***particular solution***, which explains the subscript P.

So the solution are $y = y_H + y_P$ and $y = y_P$. Furthermore, **all** the solutions to the nonhomogeneous equation are in this form.

The same methodology works for both the first-order and the second order cases. The **only** issue is how to guess the particular solution. We will see it in next section.

Here's a summary of our methods so far:

1. Rearrange the equation into the correct form. That is, put all the x-junk on the right-hand side. You should be able to reduce the equation to

$$\frac{dy}{dx} + ay = f(x)$$

for the first-order case, or

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = f(x)$$

for the second-order case

2. Using the techniques from Sections 30.4.1 and 30.4.2 above, solve the associated homogeneous equation

$$\frac{dy}{dx} + ay = 0 \quad \text{or} \quad a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0$$

The solution, which we'll write as y_H, will have one or two undetermined constants in it (depending on whether the equation is first- or second order). We call y_H the homogeneous solution of the equation

3. If the original function f is actually 0, then we're already done; the complete solution is $y = y_H$

4. On the other hand, if the function f is anything other than 0, then write down the form for the particular solution y_P. The form will have some constants which **must** be determined. **Substitute** y_P into the original equation and **equate** coefficients to find the constants

5. Finally, the solution is $y = y_H + y_P$.

● Finding a particular solution

So far, we have blissfully ignored the stuff involving x which could appear on the right-hand side (it was called $f(x)$ earlier). Now it's time to deal with it. The tactic is to write down the **form** of the particular solution, then to find the actual solution by plugging the form into the equation. The table below shows how to come up with the correct form (Page 658)

If f is a ...	then the form is ...
polynomial of degree n	$y_P =$ **general polynomial of degree n**
e.g., $f(x) = 7$	$y_P = a$
$f(x) = 3x - 2$	$y_P = ax + b$
$f(x) = 10x^2$	$y_P = ax^2 + bx + c$
$f(x) = -x^3 - x^2 + x + 22$	$y_P = ax^3 + bx^2 + cx + d$
multiple of an exponential e^{kx}	$y_P = Ce^{kx}$
e.g., $f(x) = 10e^{-4x}$	$y_P = Ce^{-4x}$
$f(x) = e^x$	$y_P = Ce^x$
multiple of $\cos(kx)$ + multiple of $\sin(kx)$	$y_P = C\cos(kx) + D\sin(kx)$
e.g., $f(x) = 2\sin(3x) - 5\cos(3x)$	$y_P = C\cos(3x) + D\sin(3x)$
$f(x) = \cos(x)$	$y_P = C\cos(x) + D\sin(x)$
$f(x) = 2\sin(11x)$	$y_P = C\cos(11x) + D\sin(11x)$
a sum or product of one of the above	**the sum or product of forms (if a product, omit a constant)**
e.g., $f(x) = 2x^2 + e^{-6x}$	$y_P = ax^2 + bx + c + Ce^{-6x}$
$f(x) = 2x^2e^{-6x}$	$y_P = (ax^2 + bx + c)e^{-6x}$
$f(x) = 7e^{2x}\sin(3x)$	$y_P = (C\cos(3x) + D\sin(3x))e^{2x}$
$f(x) = \cos(2x) + 6\sin(x)$	$y_P = C\cos(2x) + D\sin(2x) + E\cos(x) + F\sin(x)$
$f(x) = 4x\cos(3x)$	$y_P = (x + b)(C\cos(3x) + D\sin(3x))$
If y_P conflicts with y_H, multiply the form by x or x^2 as appropriate.	

This table should be fairly self-explanatory, except for the last line.

● Examples of finding particular solutions

Once you've written down the form for y_P, you still have to substitute y_P into the original differential equation in order to find the constants. To make the calculation easier, you should first calculate y_P' and y_P'' (for the first order case, you actually only need y_P') (Page 660).

● Resolving conflicts between y_P and y_H

The last line of the table in Section 30.4.5 above indicates that there might be conflicts between y_P and y_H. How can this happen? Well (Page 662)

$$y'' - 3y' + 2y = 7e^{2x}$$

$$y'' + 6y' + 9y = e^{-3x}$$

● Initial value problems (constant-coefficient linear)

Let's see how to deal with initial-value problems (IVPs) involving constant coefficient linear differential equations. As usual, to solve an IVP, first solve the differential equation, then use the initial conditions to find the remaining unknown constants

$$y' + 2y = 4x + \frac{1}{3}\sin(5x)\text{, and that } y(0) = -1$$

$$y'' - 5y' + 6y = 2x^2e^x\text{, and that } y(0) = y'(0) = 0$$

$$y'' + 6y' + 13y = 26x^3 - 3x^2 - 24x, \qquad y(0) = 1, \qquad y'(0) = 2$$

30.5 Modeling Using Differential Equations

Many quantities in the real world can be modeled (that is, theoretically approximated) by differential equations. Examples include heat flow, wave height, inflation, current in electrical circuits, and population growth, to name a few. Here's a simple example of a somewhat realistic situation involving population growth (Page 665).

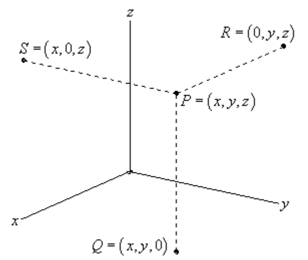
ADDITION CHAPTER 1 3-Dimensional Space

(From Paul's Online Notes)

The 3-D Coordinate System

Let's first get some basic notation out of the way. The 3-D coordinate system is often denoted by R^3. Likewise, the 2-D coordinate system is often denoted by R^2 and the 1-D coordinate system is denoted by R. Also, as you might have guessed then a general n dimensional coordinate system is often denoted by R^n.

Next, let's take a quick look at the basic coordinate system.



Collectively, the xy, xz, and yz-planes are sometimes called the coordinate planes.

Also, the point Q is often referred to as the projection of P in the xy-plane. Likewise, R is the projection of P in the yz-plane and S is the projection of P in the xz-plane.

Many of the formulas that you are used to working with in R^2 have natural extensions in R^3.

For instance, the distance between two points in R^2 is given by,

$$\mathrm{d}(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

While the distance between any two points in R^3 is given by,

$$\mathrm{d}(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Likewise, the general equation for a circle with center (h, k) and radius r is given by,

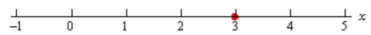
$$(x - h)^2 + (y - k)^2 = r^2$$

and the general equation for a sphere with center (h, k, l) and radius r is given by,

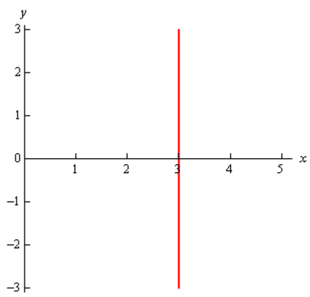
$$(x - h)^2 + (y - k)^2 + (z - l)^2 = r^2$$

Example 1: Graph $x = 3$ in R, R^2 and R^3

In R we have a single coordinate system and so $x = 3$ is a point in a 1-D coordinate system,

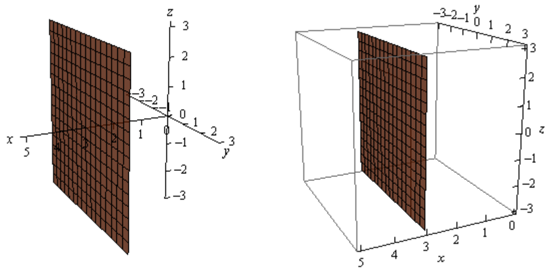


In R^2 the equation $x = 3$ tells us to graph all the points that are in the form $(3, y)$. This is a vertical line in a 2-D coordinate system,



In R^3 the equation $x = 3$ tells us to graph all the points that are in the form $(3, y, z)$. If you go back and look at the coordinate plane points this is very similar to the coordinates for the yz-plane except this time we have $x = 3$ instead of $x = 0$. So, in a 3-D coordinate system this is a plane that will be parallel to the yz-plane and pass through the x-axis at $x = 3$.

Finally, here is the graph of $x = 3$ in $\mathbb{R}^3$. Note that we've presented this graph in two different styles. On the left we've got the traditional axis system that we're used to seeing and, on the right, we've put the graph in a box. Both views can be convenient on occasion to help with perspective and so we'll often do this with 3D graphs and sketches.

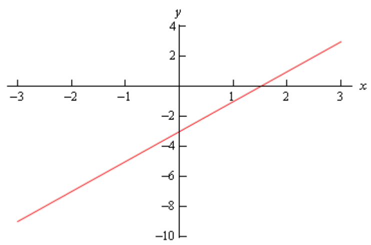


Note that at this point we can now write down the equations for each of the coordinate planes as well using this idea.

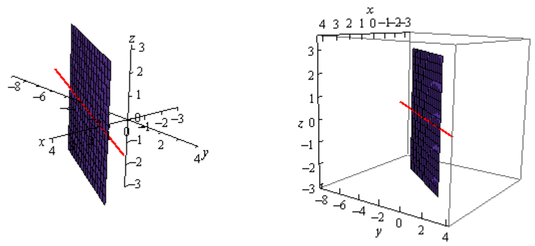
$$z = 0 \quad xy - \text{plane}$$
$$y = 0 \quad xz - \text{plane}$$
$$x = 0 \quad yz - \text{plane}$$

Example 2: Graph $y = 2x - 3$ in $\mathbb{R}^2$ and $\mathbb{R}^3$

Here is the graph in $\mathbb{R}^2$,



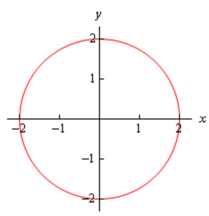
here is the graph in $\mathbb{R}^3$, because we have not specified a value of z, we are forced to let z take any value. This means that at any particular value of z we will get a copy of this line.

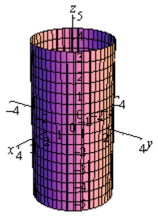


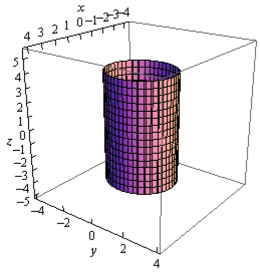
Notice that if we look to where the plane intersects the xy-plane, we will get the graph of the line in $\mathbb{R}^2$ as noted in the above graph (by the red line through the plane).

Example 3: Graph $x^2 + y^2 = 4$ in $\mathbb{R}^2$ and $\mathbb{R}^3$

Here are the graphs for this example,







The point of the examples in this section is to make sure that we are being careful with graphing equations and making sure that we always remember which coordinate system that we are in.

Equations of Lines

We're just going to need a new way of writing down the equation of a curve.

So, before we get into the equations of lines, we first need to briefly look at vector functions.

The best way to get an idea of what a vector function is to look at an example. So, consider the following vector function.

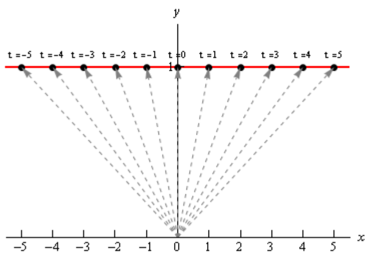
$$\vec{r}(t) = \langle t, 1 \rangle$$

A vector function is a function that takes one or more variables, one in this case, and returns a vector.

In order to find the graph of our function, we'll think of the vector that the vector function returns as a position vector for points on the graph. Recall that a position vector, say $\vec{v} = \langle a, b \rangle$, is a vector that starts at the origin and ends at the point (a, b).

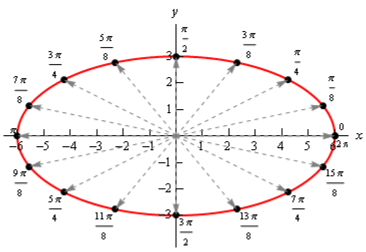
Here are some evaluations for our example,

$$\vec{r}(-3) = \langle -3,1 \rangle, \quad \vec{r}(-1) = \langle -1,1 \rangle, \quad \vec{r}(2) = \langle 2,1 \rangle, \quad \vec{r}(5) = \langle 5,1 \rangle$$



It looks like, in this case the graph of the vector equation is in fact the line $y = 1$.

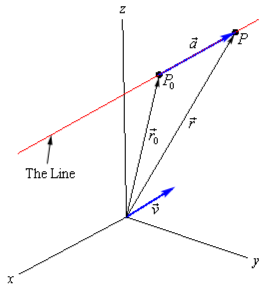
Here's another quick example. Here is the graph of $\vec{r}(t) = \langle 6\cos t, 3\sin t \rangle$,



Let's start with the following information. Suppose that we know a point that is on the line, $P_0 = (x_0, y_0, z_0)$, and that $\vec{v} = \langle a, b, c \rangle$ is some vector that is parallel to the line. Note, in all likelihood, $\vec{v}$ will not be on the line itself. We only need $\vec{v}$ to be parallel to the line. Finally, let $P = (x, y, z)$ be any point on the line.

Now, since our "slope" is a vector let's also represent the two points on the line as vectors. We'll do this with position vectors. So, let $\vec{r_0}$ and $\vec{r}$ be the position vectors for P_0 and P

respectively. Also, for no apparent reason, let's define $\vec{a}$ to be the vector with representation $\overrightarrow{P_0P}$.



since $\vec{r} - \vec{r_0} = \vec{a}$,

$$\vec{r} = \vec{r_0} + \vec{a}$$

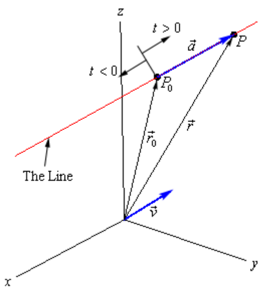
Noticing that the vectors $\vec{a}$ and $\vec{v}$ are parallel. Therefore there is a number, t, such that

$$\vec{a} = t\vec{v}$$

We now have,

$$\boxed{\vec{r} = \vec{r_0} + t\vec{v} = \langle x_0, y_0, z_0 \rangle + t\langle a, b, c \rangle}$$

This is called the **vector form of the equation of a line**. The only part of this equation that is not known is the t.



There are several other forms of the equation of a line,

$$\vec{r} = \langle x_0, y_0, z_0 \rangle + t\langle a, b, c \rangle$$

$$\langle x, y, z \rangle = \langle x_0 + ta, y_0 + tb, z_0 + tc \rangle$$

The only way for two vectors to be equal is for the components to be equal. In other words,

$$\boxed{\begin{aligned} x &= x_0 + ta \\ y &= y_0 + tb \\ z &= z_0 + tc \end{aligned}}$$

This set of equations is called the **parametric form of the equation of a line**.

There is one more form of the line that we want to look at. If we assume that a, b, and c are all non-zero numbers we can solve each of the equations in the parametric form of the line for t.

$$\boxed{\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}}$$

This is called the **symmetric equations of the line**.

If one of a, b, or c does happen to be zero we can still write down the symmetric equations. To see this let's suppose that $b = 0$,

$$\frac{x - x_0}{a} = \frac{z - z_0}{c}, \quad y = y_0$$

Example 1: Write down the equation of the line that passes through the points $(2,-1,3)$ and $(1,4,-3)$. Write down all three forms of the equation of the line,

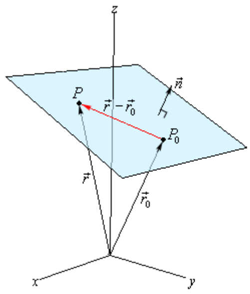
$$\vec{v} = \langle 1,-5,6 \rangle$$

$$\vec{r} = \langle 2,-1,3 \rangle + t\langle 1,-5,6 \rangle$$

$$\frac{x-2}{1} = \frac{y+1}{-5} = \frac{z-3}{6}.$$

Equations of Planes

Let's start by assuming that we know a point that is on the plane, $P_0 = (x_0, y_0, z_0)$. Let's also suppose that we have a vector that is orthogonal (perpendicular) to the plane, $\vec{n} = \langle a,b,c \rangle$. This vector is called the **normal vector**. Now, assume that $P = (x,y,z)$ is any point in the plane. Finally, since we are going to be working with vectors initially, we'll let $\vec{r_0}$ and $\vec{r}$ be the position vectors for P_0 and P respectively.



Recall from the dot product,

$$\vec{n}\cdot(\vec{r}-\vec{r_0}) = 0 \quad \Longrightarrow \quad \vec{n}\cdot\vec{r} = \vec{n}\cdot\vec{r_0}$$

This is called the **vector equation of the plane**.

A slightly more useful form of the equations is as follows. Start with the first form of the vector equation and write down a vector for the difference.

$$\langle a,b,c \rangle\cdot(\langle x,y,z \rangle - \langle x_0,y_0,z_0 \rangle) = 0$$

$$\langle a,b,c \rangle\cdot\langle x-x_0, y-y_0, z-z_0 \rangle = 0$$

Now, actually compute the dot product to get,

$$\boxed{a(x-x_0) + b(y-y_0) + c(z-z_0) = 0}$$

This is called the **scalar equation of plane**. Often this will be written as,

$$ax + by + cz = d$$

where $d = ax_0 + by_0 + cz_0$.

Notice that if we are given the equation of a plane in this form, we can quickly get a normal vector for the plane. A normal vector is,

$$\vec{n} = \langle a,b,c \rangle$$

Example 1: Determine the equation of the plane that contains the points $P = (1,-2,0)$, $Q = (3,1,4)$ and $R = (0,-1,2)$,

$$\overrightarrow{PQ} = \langle 2,3,4 \rangle \quad \text{and} \quad \overrightarrow{PR} = \langle -1,1,2 \rangle$$

using cross product,

$$\vec{n} = \overrightarrow{PQ}\times\overrightarrow{PR} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 2 & 3 & 4 \\ -1 & 1 & 2 \end{vmatrix} = 2\vec{i} - 8\vec{j} + 5\vec{k}$$

The equation of the plane is then,

$$2(x-1)-8(y+2)+5(z-0)=0$$

We used P for the point but could have used any of the three points.

Quadric Surfaces

In this section we are going to be looking at quadric surfaces. Quadric surfaces are the graphs of any equation that can be put into the general form

$$Ax^2+By^2+Cz^2+Dxy+Exz+Fyz+Gx+Hy+Iz+J=0$$

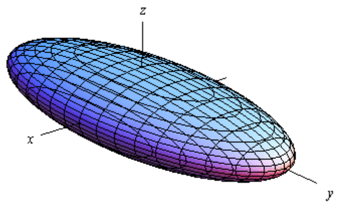
where $A,\cdots,J$ are constants.

Ellipsoid

Here is the general equation of an ellipsoid,

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$$

Here is a sketch of a typical ellipsoid,

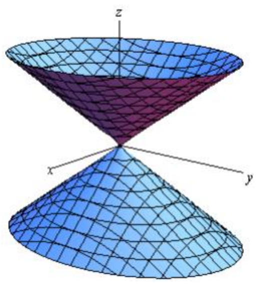


If $a=b=c$ then we will have a sphere.

Cone

Here is the general equation of a cone,

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



Now, note that while we called this a cone it is more of an hour glass shape rather than what most would call a cone. Of course, the upper and the lower portion of the hour glass really are cones as we would normally think of them,

$$z^2=c^2\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)=\frac{c^2}{a^2}x^2+\frac{c^2}{b^2}y^2=A^2x^2+B^2y^2 \quad \rightarrow \quad z=\pm\sqrt{A^2x^2+B^2y^2}$$

$z=\sqrt{A^2x^2+B^2y^2}$ will always be positive and so be the equation for just the upper portion of the "cone" above. $z=-\sqrt{A^2x^2+B^2y^2}$ will always be negative and so be the equation of just the lower portion of the "cone" above.

A cone that opens up along the x-axis will have the equation,

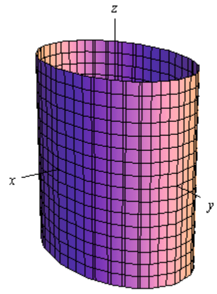
$$\frac{y^2}{b^2}+\frac{z^2}{c^2}=\frac{x^2}{a^2}$$

Cylinder

Here is the general equation of a cylinder,

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

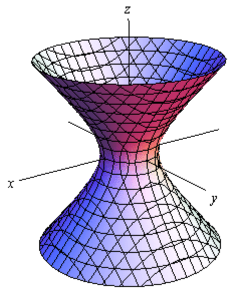
This is a cylinder whose cross section is an ellipse. If $a = b$ we have a cylinder whose cross section is a circle.



Hyperboloid of One Sheet

Here is the equation of a hyperboloid of one sheet,

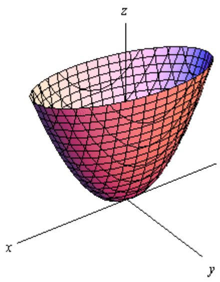
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$$



Elliptic Paraboloid

Here is the equation of an elliptic paraboloid,

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

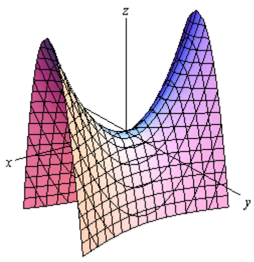


In this case the variable that isn't squared determines the axis upon which the paraboloid opens up. Also, the sign of c will determine the direction that the paraboloid opens. If c is positive then it opens up and if c is negative then it opens down.

Hyperbolic Paraboloid

Here is the equation of a hyperbolic paraboloid,

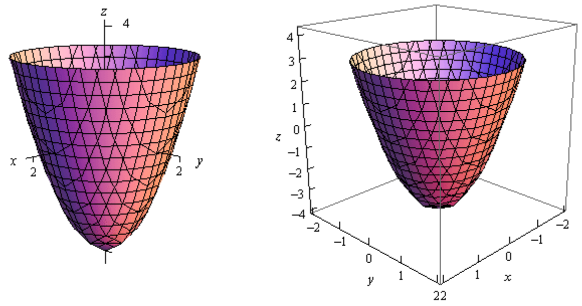
$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = \frac{z}{c}$$



These graphs are vaguely saddle shaped and the sign of c will determine the direction in which the surface "opens up".

Functions of Several Variables

First, remember that graphs of functions of two variables, $z = f(x, y)$ are surfaces in three-dimensional space. For example, here is the graph of $z = 2x^2 + 2y^2 - 4$



This is an elliptic paraboloid and is an example of a quadric surface.

Recall that the equation of a plane is given by

$$ax + by + cz = d$$

or if we solve this for z we can write it in terms of function notation. This gives,

$$f(x, y) = Ax + By + D$$

To graph a plane, we will generally find the intersection points with the three axes and then graph the triangle that connects those three points. For example,

$$f(x, y) = 12 - 3x - 4y$$

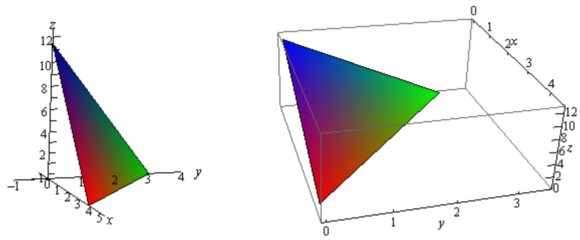
For purposes of graphing this it would probably be easier to write this as,

$$z = 12 - 3x - 4y \quad \Rightarrow \quad 3x + 4y + z = 12$$

$$x - \text{axis:} \quad (4,0,0)$$

$$y - \text{axis:} \quad (0,3,0)$$

$$z - \text{axis:} \quad (0,0,12)$$



Now, to extend this out, graphs of functions of the form $w = f(x, y, z)$ would be four-dimensional surfaces.

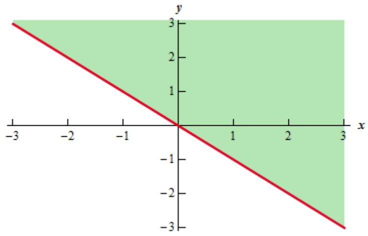
The domain of a function of a single variable is an interval (or intervals) of values from the number line, or one-dimensional space.

The domain of functions of two variables, $z = f(x, y)$, are regions from two-dimensional space and consist of all the coordinate pairs, (x, y).

Example 1: Determine the domain of each of the following,

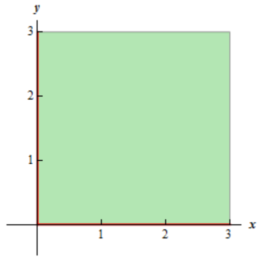
(a) $f(x, y) = \sqrt{x + y}$

$$x + y \geq 0$$



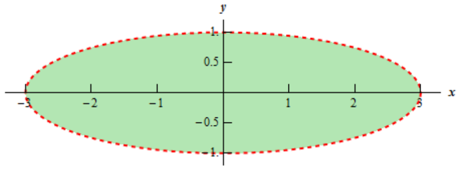
(b) $f(x,y) = \sqrt{x} + \sqrt{y}$

$$x \geq 0 \quad \text{and} \quad y \geq 0$$



(c) $f(x,y) = \ln(9 - x^2 - 9y^2)$

$$9 - x^2 - 9y^2 > 0 \quad \Rightarrow \quad \frac{x^2}{9} + y^2 < 1.$$



Example 2: Determine the domain of the following function,

$$f(x,y,z) = \frac{1}{\sqrt{x^2 + y^2 + z^2 - 16}}$$

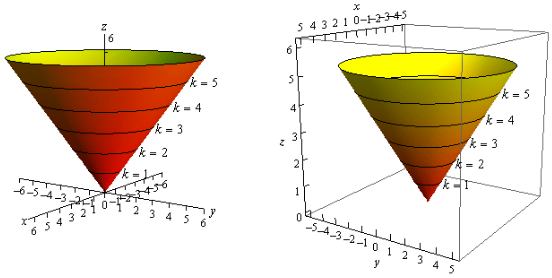
$$x^2 + y^2 + z^2 - 16 > 0 \quad \Rightarrow \quad x^2 + y^2 + z^2 > 16.$$

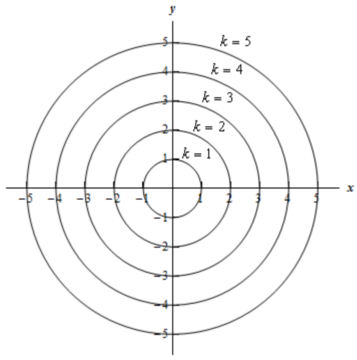
The next topic that we should look at is that of **level curves** or **contour curves**. The level curves of the function $z = f(x,y)$ are two-dimensional curves we get by setting $z = k$, where k is any number. So, the equations of the level curves are $f(x,y) = k$. Note that sometimes the equation will be in the form $f(x,y,z) = 0$ and in these cases the equations of the level curves are $f(x,y,k) = 0$.

You've probably seen level curves (or contour curves, whatever you want to call them) before. If you've ever seen the elevation map for a piece of land, this is nothing more than the contour curves for the function that gives the elevation of the land in that area.

Example 3: Identify the level curves of $f(x,y) = \sqrt{x^2 + y^2}$. Sketch a few of them,

$$k = \sqrt{x^2 + y^2}$$





Note that we can think of contours in terms of the intersection of the surface that is given by $z = f(x, y)$ and the plane $z = k$.

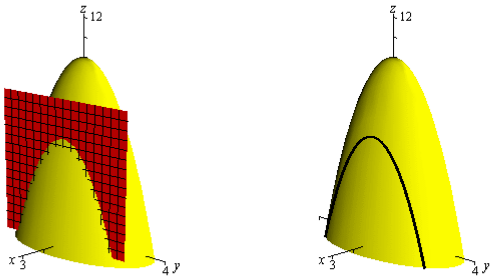
For functions of the form $f(x, y, z)$ we will occasionally look at **level surfaces**. The equations of level surfaces are given by $f(x, y, z) = k$ where k is any number.

The final topic in this section is that of **traces**. Traces of surfaces are curves that represent the intersection of the surface and the plane given by $x = a$ or $y = b$.

Example 4: Sketch the traces of $f(x, y) = 10 - 4x^2 - y^2$ for the plane $x = 1$ and $y = 2$,

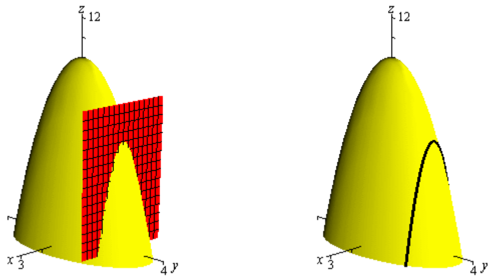
We'll start with $x = 1$,

$$z = f(1, y) = 10 - 4(1)^2 - y^2 \quad \Longrightarrow \quad z = 6 - y^2$$



For $y = 2$,

$$z = f(x, 2) = 10 - 4x^2 - (2)^2 \quad \Longrightarrow \quad z = 6 - 4x^2.$$



Vector Functions

A vector functions of a single variable in R^2 and R^3 have the form,

$$\vec{r}(t) = \langle f(t), g(t) \rangle \quad \text{and} \quad \vec{r}(t) = \langle f(t), g(t), h(t) \rangle$$

respectively, where $f(t)$, $g(t)$ and $h(t)$ are called the *component functions*.

The *domain* of a vector function is the set of all t's for which all the component functions are defined.

Example 1: Determine the domain of the following function,

$$\vec{r}(t) = \langle \cos t, \ln(4 - t), \sqrt{t + 1} \rangle$$

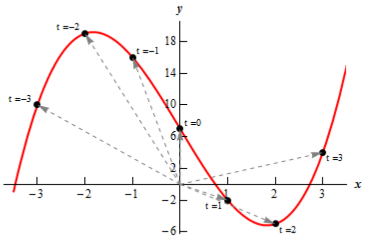
The domain is $t < 4$ and $t \geq -1$,

$$[-1,4).$$

In order to graph a vector function, all we do is to think of the vector returned by the vector function as a position vector for points on the graph.

So, in order to sketch the graph of a vector function, all we need to do is to plug in some values of t and then plot points that correspond to the resulting position vector we get out of the vector function.

Example 2: Sketch the graph of $\vec{r}(t) = \langle t, t^3 - 10t + 7 \rangle$,



Any vector function can be broken down into a set of parametric equations that represent the same graph. In general, the two-dimensional vector function, $\vec{r}(t) = \langle f(t), g(t) \rangle$, can be broken down into the parametric equations,

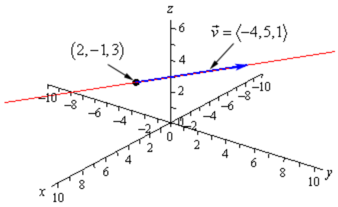
$$x = f(t) \quad \text{and} \quad y = g(t)$$

Likewise, a three-dimensional vector function, $\vec{r}(t) = \langle f(t), g(t), h(t) \rangle$, can be broken down into the parametric equations,

$$x = f(t) \quad y = g(t) \quad \text{and} \quad z = h(t).$$

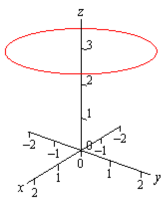
Example 4: Sketch the graph of the following vector function,

$$\vec{r}(t) = \langle 2 - 4t, -1 + 5t, 3 + t \rangle$$
$$\vec{r}(t) = \langle 2, -1, 3 \rangle + t\langle -4, 5, 1 \rangle$$



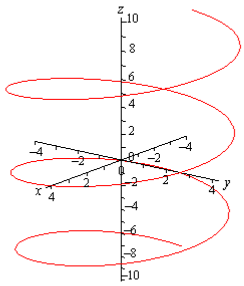
Example 5: Sketch the graph of the following vector function,

$$\vec{r}(t) = \langle 2\cos t, 2\sin t, 3 \rangle$$
$$x = 2\cos t \quad y = 2\sin t \quad \text{and} \quad z = 3$$



Example 6: Sketch the graph of the following vector function,

$$\vec{r}(t) = \langle 4\cos t, 4\sin t, t \rangle$$



So, we've got a **helix** (or spiral, depending on what you want to call it) here.
As with circles the component that has the t will determine the axis that the helix rotates about.
For instance,

$$\vec{r}(t) = \langle t, 6\cos t, 6\sin t\rangle$$

is a helix that rotates around the x-axis.
Also note that if we allow the coefficients on the sine and cosine for both the circle and helix to be different, we will get ellipses. For example,

$$\vec{r}(t) = \langle 9\cos t, t, 2\sin t\rangle$$

will be a helix that rotates about the y-axis and is in the shape of an ellipse.
Example 7: Determine the vector equation for the line segment starting at the point $P = (x_1, y_1, z_1)$ and ending at the point $Q = (x_2, y_2, z_2)$,

$$\vec{r}(t) = \langle x_1, y_1, z_1\rangle + t\langle x_2 - x_1, y_2 - y_1, z_2 - z_1\rangle$$

$$\vec{r}(t) = (1 - t)\langle x_1, y_1, z_1\rangle + t\langle x_2, y_2, z_2\rangle \quad 0 \le t \le 1$$

The graphs of vector function of two variables are surfaces.
Example 8: Identify the surface that is described by $\vec{r}(x, y) = x\vec{\imath} + y\vec{\jmath} + (x^2 + y^2)\vec{k}$,

$$x = x \quad y = y \quad \text{and} \quad z = x^2 + y^2$$

The first two are really only acknowledging that we are picking x and y for free. The third equation is the equation of an elliptic paraboloid and so the vector function represents an elliptic paraboloid.
As a final topic for this section let's generalize the idea from the previous example. Note that given any function of one variable ($y = f(x)$ or $x = h(x)$) or any function of two variables ($z = g(x, y)$, $x = g(y, z)$, or $y = g(x, z)$), we can always write down a vector form of the equation.
For a function of one variable this will be,

$$\vec{r}(x) = x\vec{\imath} + f(x)\vec{\jmath} \quad \text{and} \quad \vec{r}(y) = h(y)\vec{\imath} + y\vec{\jmath}$$

and for a function of two variables the vector form will be,

$$\vec{r}(x, y) = x\vec{\imath} + y\vec{\jmath} + g(x, y)\vec{k} \quad \text{and} \quad \vec{r}(y, z) = g(y, z)\vec{\imath} + y\vec{\jmath} + z\vec{k}$$

$$\vec{r}(x, z) = x\vec{\imath} + g(x, z)\vec{\jmath} + z\vec{k}$$

depending upon the original form of the function.
Calculus with Vector Functions
We will be doing all of the work in R^3 but we can naturally extend the formulas/work in this section to R^n (i.e. n-dimensional space).
Here is the limit of a vector function,

$$\lim_{t\to a} \vec{r}(t) = \lim_{t\to a}\langle f(t), g(t), h(t)\rangle$$

$$= \langle \lim_{t\to a} f(t), \lim_{t\to a} g(t), \lim_{t\to a} h(t)\rangle$$

$$= \lim_{t\to a} f(t)\,\vec{\imath} + \lim_{t\to a} g(t)\,\vec{\jmath} + \lim_{t\to a} h(t)\,\vec{k}$$

Now let's take care of derivatives,

$$\vec{r}\,'(t) = \langle f'(t), g'(t), h'(t)\rangle = f'(t)\vec{\imath} + g'(t)\vec{\jmath} + h'(t)\vec{k}$$

Most of the basic facts that we know about derivatives still hold however, just to make it clear here are some facts about derivatives of vector functions.

$$\frac{d}{dt}(\vec{u} + \vec{v}) = \vec{u}\,' + \vec{v}\,'$$

$$(c\vec{u})' = c\vec{u}\,'$$

$$\frac{d}{dt}\big(f(t)\vec{u}(t)\big) = f'(t)\vec{u}(t) + f(t)\vec{u}\,'$$

$$\frac{d}{dt}(\vec{u}\cdot\vec{v}) = \vec{u}\,'\cdot\vec{v} + \vec{u}\cdot\vec{v}\,'$$

$$\frac{d}{dt}(\vec{u}\times\vec{v}) = \vec{u}\,'\times\vec{v} + \vec{u}\times\vec{v}\,'$$

$$\frac{d}{dt}\Big(\vec{u}\big(f(t)\big)\Big) = f'(t)\vec{u}\,'\big(f(t)\big)$$

A **smooth curve** is any curve for which $\vec{r}\,'(t)$ is continuous and $\vec{r}\,'(t) \neq \vec{0}$ for any t except possibly at the endpoints. A helix is a smooth curve, for example.

For indefinite integrals,

$$\int \vec{r}(t)dt = \langle \int f(t)dt, \int g(t)dt, \int h(t)dt\rangle + \vec{c}$$

$$\int \vec{r}(t)dt = \int f(t)dt\,\vec{\imath} + \int g(t)dt\,\vec{\jmath} + \int h(t)dt\,\vec{k} + \vec{c}$$

and the following for definite integrals,

$$\int_a^b \vec{r}(t)dt = \left(\langle \int f(t)dt, \int g(t)dt, \int h(t)dt\rangle\right)\Big|_a^b$$

$$\int_a^b \vec{r}(t)dt = \left(\int f(t)dt\,\vec{\imath} + \int g(t)dt\,\vec{\jmath} + \int h(t)dt\,\vec{k}\right)\Big|_a^b$$

With the indefinite integrals, we put in a constant of integration to make sure that, it was clear that the constant in this case needs to be a vector instead of a regular constant.

Example 3: Compute $\int \vec{r}(t)dt$ for $\vec{r}(t) = \langle \sin(t), 6, 4t\rangle$,

$$\int \vec{r}(t)dt = \langle -\cos(t), 6t, 2t^2\rangle + \vec{c}.$$

Example 4: Compute $\int_0^1 \vec{r}(t)dt$ for $\vec{r}(t) = \langle \sin(t), 6, 4t\rangle$,

$$\int_0^1 \vec{r}(t)dt = (\langle -\cos(t), 6t, 2t^2\rangle)\Big|_0^1$$
$$= \langle -\cos(1), 6,2\rangle - \langle -1,0,0\rangle$$
$$= \langle 1-\cos(1), 6,2\rangle.$$

Tangent, Normal and Binormal Vectors

With vector functions we get exactly the same result, with one exception. Given the vector function, $\vec{r}(t)$, we call $\vec{r}'(t)$ the **tangent vector** provided it exists and provided $\vec{r}'(t) \neq \vec{0}$. The tangent line to $\vec{r}(t)$ at P is then the line that passes through the point P and is parallel to the tangent vector, $\vec{r}'(t)$. Note that we really do need to require $\vec{r}'(t) \neq \vec{0}$ in order to have a tangent vector. If we had

$$\vec{r}'(t) = \vec{0}$$

we would have a vector that had no magnitude, so it couldn't give us the direction of the tangent.

Also, provided $\vec{r}'(t) \neq \vec{0}$, the **unit tangent vector** to the curve is given by,

$$\boxed{\vec{T}(t) = \frac{\vec{r}'(t)}{\|\vec{r}'(t)\|}}$$

Example 1: Find the general formula for the tangent vector and unit tangent vector to the curve given by $\vec{r}(t) = t^2\vec{\imath} + 2\sin t\,\vec{\jmath} + 2\cos t\,\vec{k}$,

$$\vec{r}'(t) = 2t\vec{\imath} + 2\cos t\,\vec{\jmath} - 2\sin t\,\vec{k}$$
$$\|\vec{r}'(t)\| = \sqrt{4t^2+4}$$
$$\vec{T}(t) = \frac{1}{\sqrt{4t^2+4}}\left(2t\vec{\imath} + 2\cos t\,\vec{\jmath} - 2\sin t\,\vec{k}\right).$$

Example 2: Find the vector equation of the tangent line to the curve given by $\vec{r}(t) = t^2\vec{\imath} + 2\sin t\,\vec{\jmath} + 2\cos t\,\vec{k}$ at $t = \frac{\pi}{3}$,

The tangent vector,

$$\vec{r}'\left(\frac{\pi}{3}\right) = \frac{2\pi}{3}\vec{\imath} + \vec{\jmath} - \sqrt{3}\vec{k}$$

The point on the line at $t = \frac{\pi}{3}$ so,

$$\vec{r}\left(\frac{\pi}{3}\right) = \frac{\pi^2}{9}\vec{\imath} + \sqrt{3}\vec{\jmath} + \vec{k}$$

The vector equation of the line is then,

$$\vec{v}(t) = \langle\frac{\pi^2}{9}, \sqrt{3}, 1\rangle + t\,\langle\frac{2\pi}{3}, 1, -\sqrt{3}\rangle.$$

Next, we need to talk about the **unit normal** and the **binormal** vectors.

The unit normal vector is defined to be,

$$\boxed{\vec{N}(t) = \frac{\vec{T}'(t)}{\|\vec{T}'(t)\|}}$$

The unit normal is orthogonal (or normal, or perpendicular) to the unit tangent vector and hence to the curve as well.

The definition of the unit normal vector always seems a little mysterious when you first see it. It follows directly from the following fact.

Fact

Suppose that $\vec{r}(t)$ is a vector such that $\|\vec{r}(t)\| = c$ for all t. Then $\vec{r}\,'(t)$ is orthogonal to $\vec{r}(t)$.

To prove this fact is pretty simple. From the fact statement and the relationship between the magnitude of a vector and the dot product, we have the following,

$$\vec{r}(t)\cdot\vec{r}(t) = \|\vec{r}(t)\|^2 = c^2 \quad \text{for all } t$$

Now, because this is true for all t we can see that,

$$\frac{d}{dt}\big(\vec{r}(t)\cdot\vec{r}(t)\big) = \vec{r}\,'(t)\cdot\vec{r}(t) + \vec{r}(t)\cdot\vec{r}\,'(t) = 2\vec{r}\,'(t)\cdot\vec{r}(t)$$

Or, upon putting all this together we get,

$$2\vec{r}\,'(t)\cdot\vec{r}(t) = 0 \quad \Longrightarrow \quad \vec{r}\,'(t)\cdot\vec{r}(t) = 0$$

Therefore $\vec{r}\,'(t)$ is orthogonal to $\vec{r}(t)$.

The definition of the unit normal then falls directly from this. Because $\vec{T}(t)$ is a unit vector we know that $\|\vec{T}(t)\| = 1$ for all t. Hence, $\vec{T}\,'(t)$ is orthogonal to $\vec{T}(t)$.

Next, is the binormal vector. The binormal vector is defined to be,

$$\boxed{\vec{B}(t) = \vec{T}(t)\times\vec{N}(t)}$$

The binormal vector is orthogonal to both the tangent vector and the normal vector.

Example 3: Find the normal and binormal vectors for $\vec{r}(t) = \langle t, 3\sin t, 3\cos t\rangle$,

$$\vec{r}\,'(t) = \langle 1,3\cos t, -3\sin t\rangle$$

$$\|\vec{r}\,'(t)\| = \sqrt{1 + 9\cos^2 t + 9\sin^2 t} = \sqrt{10}$$

$$\vec{T}(t) = \langle \frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\cos t, -\frac{3}{\sqrt{10}}\sin t\rangle$$

$$\vec{T}\,'(t) = \langle 0, -\frac{3}{\sqrt{10}}\sin t, -\frac{3}{\sqrt{10}}\cos t\rangle$$

$$\|\vec{T}\,'(t)\| = \frac{3}{\sqrt{10}}$$

$$\vec{N}(t) = \frac{\sqrt{10}}{3}\langle 0, -\frac{3}{\sqrt{10}}\sin t, -\frac{3}{\sqrt{10}}\cos t\rangle = \langle 0, -\sin t, -\cos t\rangle$$

$$\vec{B}(t) = \vec{T}(t)\times\vec{N}(t)$$

$$= -\frac{3}{\sqrt{10}}\vec{i} + \frac{1}{\sqrt{10}}\cos t\,\vec{j} - \frac{1}{\sqrt{10}}\sin t\,\vec{k}.$$

Arc Length with Vector Functions

We want to determine the length of a vector function,

$$\vec{r}(t) = \langle f(t), g(t), h(t)\rangle$$

on the interval $a \le t \le b$.

Recall that we can write the vector function into the parametric form,

$$x = f(t), \quad y = g(t) \quad \text{and} \quad z = h(t)$$

Also, recall that with two dimensional parametric curves the arc length is given by,

$$L = \int_a^b \sqrt{[f'(t)]^2 + [g'(t)]^2}\,dt$$

There is a natural extension of this to three dimensions. So, the length of the curve $\vec{r}(t)$ on the interval $a \le t \le b$ is,

$$L = \int_a^b \sqrt{[f'(t)]^2 + [g'(t)]^2 + [h'(t)]^2}\,dt$$

Notice that the integrand (the function we're integrating) is nothing more than the magnitude of the tangent vector,

$$\|\vec{r}'(t)\| = \sqrt{[f'(t)]^2 + [g'(t)]^2 + [h'(t)]^2}$$

Therefore, the arc length can be written as,

$$\boxed{L = \int_a^b \|\vec{r}'(t)\|\,dt}$$

We need to take a quick look at another concept here. We define the **arc length function** as,

$$\boxed{s(t) = \int_0^t \|\vec{r}'(u)\|\,du}$$

Before we look at why this might be important let's work a quick example,

Example 2: Determine the arc length function for $\vec{r}(t) = \langle 2t, 3\sin(2t), 3\cos(2t)\rangle$,

$$\|\vec{r}'(t)\| = 2\sqrt{10}$$

$$s(t) = \int_0^t 2\sqrt{10}\,du = 2\sqrt{10}t.$$

Okay, just why would we want to do this? Well let's take the result of the example above and solve it for t,

$$t = \frac{s}{2\sqrt{10}}$$

Now, taking this and plugging it into the original vector function and we can **reparametrize** the function into the form, $\vec{r}(t(s))$. For our function this is,

$$\vec{r}(t(s)) = \left\langle \frac{s}{\sqrt{10}}, 3\sin\left(\frac{s}{\sqrt{10}}\right), 3\cos\left(\frac{s}{\sqrt{10}}\right)\right\rangle$$

So, why would we want to do this? Well with the reparameterization we can now tell where we are on the curve after we've traveled a distance of s along the curve. Note as well that we will start the measurement of distance from where we are at $t = 0$.

Curvature

In this section we want to briefly discuss the **curvature** of a smooth curve. ($\vec{r}'(t)$ is continuous and $\vec{r}'(t) \neq 0$) The curvature measures how fast a curve is changing direction at a given point. There are several formulas for determining the curvature for a curve. The formal definition of curvature is,

$$\kappa = \left|\frac{d\vec{T}}{ds}\right|$$

where $\vec{T}$ is the unit tangent and s is the arc length. (Reparametrize a curve to get it into terms of the arc length)

In general, the formal definition of the curvature is not easy to use so there are two alternate formulas that we can use. Here they are. (no proof)

$$\kappa = \frac{\|\vec{T}'(t)\|}{\|\vec{r}'(t)\|} \quad \text{or} \quad \kappa = \frac{\|\vec{r}'(t) \times \vec{r}''(t)\|}{\|\vec{r}'(t)\|^3}$$

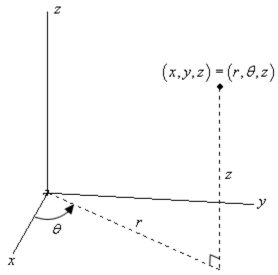
Velocity and Acceleration

Cylindrical Coordinates

As with two-dimensional space the standard (x, y, z) coordinate system is called the Cartesian coordinate system. In the last two sections of this chapter we'll be looking at some alternate coordinate systems for three-dimensional space.

We'll start off with the cylindrical coordinate system. This one is fairly simple as it is nothing more than an extension of **polar coordinates** into three dimensions. Not only is it an extension of polar coordinates, but we extend it into the third dimension just as we extend Cartesian coordinates into the third dimension. All that we do is add a z on as the third coordinate. The r and θ are the same as with polar coordinates.

Here is a sketch of a point in $\mathbb{R}^3$.



The conversions for x and y are the same conversions that we used back when we were looking at polar coordinates. So, if we have a point in cylindrical coordinates, the Cartesian coordinates can be found by using the following conversions.

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= z \end{aligned}$$

The third equation is just an acknowledgement that the z-coordinate of a point in Cartesian and polar coordinates is the same.

Likewise, if we have a point in Cartesian coordinates the cylindrical coordinates can be found by using the following conversions.

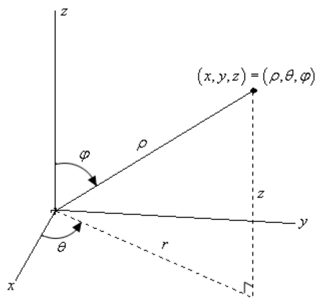
$$r = \sqrt{x^2 + y^2} \quad \text{OR} \quad r^2 = x^2 + y^2$$

$$\theta = \tan^{-1}\left(\frac{y}{x}\right)$$

$$z = z$$

Spherical Coordinates

In this section we will introduce spherical coordinates. Spherical coordinates can take a little getting used to. It's probably easiest to start things off with a sketch.



Spherical coordinates consist of the following three quantities.

First there is ρ. This is the distance from the origin to the point and we will require $\rho \geq 0$.

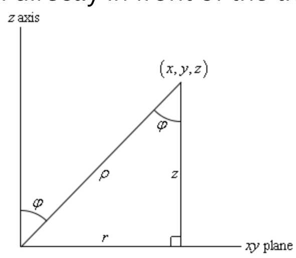
Next there is θ. This is the same angle that we saw in polar/cylindrical coordinates. It is the angle between the positive x-axis and the line above denoted by r (which is also the same r as in polar/cylindrical coordinates). There are no restrictions on θ.

Finally, there is φ. This is the angle between the positive z-axis and the line from the origin to the point. We will require $0 \leq \varphi \leq \pi$.

In summary, ρ is the distance from the origin to the point, φ is the angle that we need to rotate down from the positive z-axis to get to the point, and θ is how much we need to rotate around the z-axis to get to the point.

We should first derive some conversion formulas. Let's **first** start with a point in spherical coordinates and ask what the cylindrical coordinates of the point are. So, we know (ρ, θ, φ) and want to find (r, θ, z).

If we look at the sketch above from directly in front of the triangle, we get the following sketch,



$$\boxed{\begin{aligned} r &= \rho \sin\varphi \\ \theta &= \theta \\ z &= \rho \cos\varphi \end{aligned}}$$

Note as well from the Pythagorean theorem we also get,

$$\boxed{\rho^2 = r^2 + z^2}$$

Next, let's find the Cartesian coordinates of the same point. To do this we'll start with the cylindrical conversion formulas from the previous section.

$$x = r\cos\theta$$
$$y = r\sin\theta$$
$$z = z$$

$$\boxed{\begin{aligned} x &= \rho \sin\varphi \cos\theta \\ y &= \rho \sin\varphi \sin\theta \\ z &= \rho \cos\varphi \end{aligned}}$$

Also note that since we know that $r^2 = x^2 + y^2$ we get

$$\boxed{\rho^2 = x^2 + y^2 + z^2}$$

Converting points from Cartesian or cylindrical coordinates into spherical coordinates is usually done with the same conversion formulas.

ADDITION CHAPTER 2 Partial Derivatives

Partial Derivatives

Limits

(From Paul's Online Notes)

In this section we will take a look at limits involving functions of more than one variable. In fact, we will concentrate mostly on limits of functions of two variables, but the ideas can be extended out to functions with more than two variables.

Before getting into this, let's briefly recall how limits of functions of one variable work. We say that,

$$\lim_{x \to a} f(x) = L$$

provided,

$$\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = L$$

we will have $\lim_{x \to a} f(x) = L$ provided $f(x)$ approaches L as we move in towards $x = a$ (without letting $x = a$) from both sides.

Now, notice that in this case there are only two paths that we can take as we move in towards $x = a$. We can either move in from the left or we can move in from the right. Then in order for the limit of a function of one variable to exist the function must be approaching the same value as we take each of these paths in towards $x = a$.

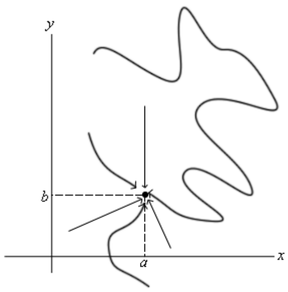
With functions of **two variables** we will have to do something similar, except this time there is (potentially) going to be a lot more work involved. Let's first address the notation and get a feel for just what we're going to be asking for in these kinds of limits.

We will be asking to take the limit of the function $f(x, y)$ as x approaches a and as y approaches b. This can be written in several ways. Here are a couple of the more standard notations.

$$\lim_{\substack{x \to a \\ y \to b}} f(x, y) \quad \text{or} \quad \lim_{(x,y) \to (a,b)} f(x, y)$$

We will use the second notation more often than not in this course. The second notation is also a little more helpful in illustrating what we are really doing here when we are taking a limit. In taking a limit of a function of two variables, we are really asking what the value of $f(x, y)$ is doing as we move the point (x, y) in closer and closer to the point (a, b) without actually letting it be (a, b).

Just like with limits of functions of one variable, in order for this limit to exist, the function must be approaching the same value regardless of the path that we take as we move in towards (a, b). The problem that we are immediately faced with is that there are literally an infinite number of paths that we can take as we move in towards (a, b). Here are a few examples of paths that we could take.



To show that a limit exists, we would technically need to check an infinite number of paths and verify that the function is approaching the same value regardless of the path we are using to approach the point.

Luckily for us however we can use one of the main ideas from Calculus I limits to help us take limits here.

Definition

A function $f(x,y)$ is continuous at the point (a,b) if,

$$\lim_{(x,y)\to(a,b)} f(x,y) = f(a,b)$$

From a graphical standpoint, this definition means the same thing as it did when we first saw continuity in Calculus I. A function will be continuous at a point if the graph doesn't have any holes or breaks at that point.

How can this help us take limits? Well, just as in Calculus I, if you know that a function is continuous at (a,b), then you also know that

$$\lim_{(x,y)\to(a,b)} f(x,y) = f(a,b)$$

must be true. So, if we know that a function is continuous at a point, then all we need to do to take the limit of the function at that point is to plug the point into the function.

All the standard functions that we know to be continuous are **still** continuous even if we are plugging in more than one variable now. We just need to watch out for division by zero, square roots of negative numbers, logarithms of zero or negative numbers, etc.

Note that the idea about paths is one that we shouldn't forget, since it is a nice way to determine if a limit doesn't exist. If we can find two paths upon which the function approaches different values as we get near the point, then we will know that the limit doesn't exist.

Let's take a look at a couple of examples.

Example 1: Determine if the following limits exist or not. If they do exist give the value of the limit.

(a) $\displaystyle\lim_{(x,y,z)\to(2,1,-1)} 3x^2z + yx\cos(\pi x - \pi z)$

Okay, in this case the function is continuous at the point in question and so all we need to do is plug in the values and we're done.

$$\lim_{(x,y,z)\to(2,1,-1)} 3x^2z + yx\cos(\pi x - \pi z) = 3(2)^2(-1) + (1)(2)\cos(2\pi + \pi) = -14$$

(b) $\displaystyle\lim_{(x,y)\to(5,1)} \frac{xy}{x+y}$

In this case the function will not be continuous along the line $y = -x$, since we will get division by zero when this is true. However, for this problem that is not something that we will need to worry about since the point that we are taking the limit at isn't on this line.
Therefore, all that we need to do is plug in the point since the function is continuous at this point.

$$\lim_{(x,y)\to(5,1)} \frac{xy}{x+y} = \frac{5}{6}$$

In the previous example there wasn't really anything to the limits. Let's work a few examples that are more typical of those you'll see here.
Example 2: Determine if the following limit exist or not. If they do exist give the value of the limit.

$$\lim_{(x,y)\to(1,1)} \frac{2x^2 - xy - y^2}{x^2 - y^2}$$

In this case the function is not continuous at the point in question (clearly division by zero). However, that does not mean that the limit can't be done. We saw many examples of this in Calculus I where the function was not continuous at the point we were looking at, and yet the limit did exist.
In the case of this limit, notice that we can factor both the numerator and denominator of the function as follows,

$$\lim_{(x,y)\to(1,1)} \frac{2x^2 - xy - y^2}{x^2 - y^2} = \lim_{(x,y)\to(1,1)} \frac{(2x+y)(x-y)}{(x-y)(x+y)} = \lim_{(x,y)\to(1,1)} \frac{2x+y}{x+y}$$

So, upon factoring and canceling common factors, we arrive at a function that in fact we can take the limit of. So, to finish out this example, all we need to do is actually take the limit.
Taking the limit gives,

$$\lim_{(x,y)\to(1,1)} \frac{2x^2 - xy - y^2}{x^2 - y^2} = \lim_{(x,y)\to(1,1)} \frac{2x+y}{x+y} = \frac{3}{2}$$

However, previous example tends to be the exception in the examples/problems as the next set of examples will show.
Example 3: Determine if the following limits exist or not. If they do exist give the value of the limit.

(a) $\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{x^4+3y^4}$

In this case the function is not continuous at the point in question and so we can't just plug in the point. Also, note that, unlike the previous example, we can't factor this function and do some canceling so that the limit can be taken.
Therefore, since the function is not continuous at the point and because there is no factoring we can do, there is at least a chance that the limit doesn't exist. If we could find two different paths to approach the point that gave different values for the limit, then we would know that the limit didn't exist. Two of the more common paths to check are the x and y-axis, so let's try those.
When we approach a point along a path, we will do this by either fixing x or y or by relating x and y through some function. In this way we can reduce the limit to just a limit, involving a single variable which we know how to do from Calculus I.
So, let's see what happens along the x-axis.

$$\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{x^4+3y^4} = \lim_{(x,0)\to(0,0)} \frac{x^2(0)^2}{x^4+3(0)^4} = \lim_{(x,0)\to(0,0)} 0 = 0$$

So, along the x-axis the function will approach zero as we move in towards the origin.

Now, let's try the y-axis.

$$\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{x^4+3y^4} = \lim_{(0,y)\to(0,0)} \frac{(0)^2y^2}{(0)^4+3y^4} = \lim_{(0,y)\to(0,0)} 0 = 0$$

So, the same limit along two paths. Don't misread this. This does NOT say that the limit exists and has a value of zero. This only means that the limit happens to have the same value along two paths.

Let's take a look at a third fairly common path to take a look at. In this case we'll move in towards the origin along the path $y = x$.

$$\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{x^4+3y^4} = \lim_{(x,x)\to(0,0)} \frac{x^2x^2}{x^4+3x^4} = \lim_{(x,x)\to(0,0)} \frac{x^4}{4x^4} = \lim_{(x,x)\to(0,0)} \frac{1}{4} = \frac{1}{4}$$

So, a different value from the previous two paths and this means that the limit can't possibly exist.

Note that we can use this idea of moving in towards the origin along a line with the more general path $y = mx$ if we need to.

(b) $\lim_{(x,y)\to(0,0)} \frac{x^3y}{x^6+y^2}$

Okay, with this last one we again have continuity problems at the origin, and again there is no factoring we can do that will allow the limit to be taken. So, again let's see if we can find a couple of paths that give different values of the limit.

First, we will use the path $y = x$. Along this path we have,

$$\lim_{(x,y)\to(0,0)} \frac{x^3y}{x^6+y^2} = \lim_{(x,x)\to(0,0)} \frac{x^3x}{x^6+x^2} = \lim_{(x,x)\to(0,0)} \frac{x^4}{x^6+x^2} = \lim_{(x,x)\to(0,0)} \frac{x^2}{x^4+1} = 0$$

Now, let's try the path $y = x^3$. Along this path the limit becomes,

$$\lim_{(x,y)\to(0,0)} \frac{x^3y}{x^6+y^2} = \lim_{(x,x^3)\to(0,0)} \frac{x^3x^3}{x^6+(x^3)^2} = \lim_{(x,x^3)\to(0,0)} \frac{x^6}{2x^6} = \lim_{(x,x^3)\to(0,0)} \frac{1}{2} = \frac{1}{2}$$

We now have two paths that give different values for the limit and so the limit doesn't exist.

Partial Derivatives

Now that we have the brief discussion on limits. Out of the way we can proceed into taking derivatives of functions of more than one variable.

Recall that given a function of one variable, $f(x)$, the derivative, $f'(x)$, represents the rate of change of the function as x changes.

The problem with functions of more than one variable is that there is more than one variable. In other words, what do we do if we only want one of the variables to change, or if we want more than one of them to change? In fact, if we're going to allow more than one of the variables to change, there are then going to be an infinite amount of ways for them to change.

We will need to develop ways, and notations, for dealing with all of these cases. In this section we are going to concentrate exclusively on only changing one of the variables at a time, while the remaining variable(s) are held fixed.

Let's start with the function $f(x, y) = 2x^2y^2$ and let's determine the rate at which the function is changing at a point, (a, b), if we hold y fixed and allow x to vary and if we hold x fixed and allow y to vary.

We'll start by looking at the case of holding y fixed and allowing x to vary. Doing this will give us a function involving only x's and we can define a new function as follows,

$$g(x) = f(x,b) = 2x^2b^3$$

Here is the rate of change of the function at (a,b) if we hold y fixed and allow x to vary.

$$g'(a) = 4ab^3$$

We will call $g'(a)$ the **partial derivative** of $f(x,y)$ with respect to x at (a,b) and we will denote it in the following way,

$$f_x(a,b) = 4ab^3$$

Now, let's do it the other way. We will now hold x fixed and allow y to vary.

$$h(x) = f(a,y) = 2a^2y^3 \quad \Rightarrow \quad h'(b) = 6a^2b^2$$

In this case we call $h'(b)$ the **partial derivative** of $f(x,y)$ with respect to y at (a,b) and we denote it as follows,

$$f_y(a,b) = 6a^2b^2$$

Note that these two partial derivatives are sometimes called the **first order partial derivatives**. Just as with functions of one variable, we can have derivatives of all orders.

Note as well that we usually don't use the (a,b) notation for partial derivatives, as that implies we are working with a specific point which we usually are not doing. The more standard notation is to just continue to use (x,y). So, the partial derivatives from above will more commonly be written as,

$$f_x(x,y) = 4xy^3 \quad \text{and} \quad f_y(x,y) = 6x^2y^2$$

Now, as this quick example has shown, taking derivatives of functions of more than one variable is done in pretty much the same manner as taking derivatives of a single variable. To compute $f_x(x,y)$, all we need to do is treat all the y's as constants (or numbers) and then differentiate the x's as we've always done. Likewise, to compute $f_y(x,y)$ we will treat all the x's as constants and then differentiate the y's as we are used to doing.

Before we work any examples, let's get the formal definition of the partial derivative out of the way as well as some alternate notation.

$$f_x(x,y) = \lim_{h\to 0}\frac{f(x+h,y)-f(x,y)}{h} \quad \text{and} \quad f_y(x,y) = \lim_{h\to 0}\frac{f(x,y+h)-f(x,y)}{h}$$

Now let's take a quick look at some of the possible alternate notations for partial derivatives. Given the function $z = f(x,y)$ the following are all equivalent notations,

$$f_x(x,y) = f_x = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\big(f(x,y)\big) = z_x = \frac{\partial z}{\partial x} = D_x f$$

$$f_y(x,y) = f_y = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y}\big(f(x,y)\big) = z_y = \frac{\partial z}{\partial y} = D_y f$$

For the fractional notation of the partial derivative, notice the difference between the partial derivative and the ordinary derivative from single variable calculus.

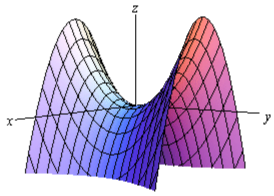
$$f(x) \quad \Rightarrow \quad f'(x) = \frac{df}{dx}$$

$$f(x,y) \quad \Rightarrow \quad f_x(x,y) = \frac{\partial f}{\partial x} \text{ and } f_y(x,y) = \frac{\partial f}{\partial y}$$

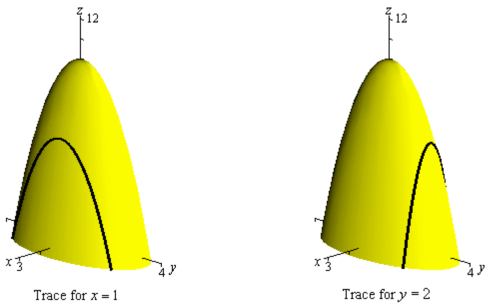
Implicit differentiation works in exactly the same manner with functions of multiple variables.

Interpretations of Partial Derivatives

As we saw in the previous section, $f_x(x,y)$ represents the rate of change of the function $f(x,y)$ as we change x and hold y fixed, while $f_y(x,y)$ represents the rate of change of $f(x,y)$ as we change y and hold x fixed.
Note that it is completely possible for a function to be increasing for a fixed y and decreasing for a fixed x at a point. As the following figure shows,



This is a graph of a hyperbolic paraboloid, and at the origin we can see that if we move in along the y-axis, the graph is increasing and if we move along the x-axis the graph is decreasing.
The next interpretation was one of the standard interpretations in a Calculus I class. We know from a Calculus I class that $f'(a)$ represents the slope of the tangent line to $y = f(x)$ at $x = a$. Well, $f_x(a,b)$ and $f_y(a,b)$ also represent the slopes of tangent lines. The difference here is the functions that they represent tangent lines to.
Partial derivatives are the slopes of **traces**. The partial derivative $f_x(a,b)$ is the slope of the trace of $f(x,y)$ for the plane $y = b$ at the point (a,b). Likewise, the partial derivative $f_y(a,b)$ is the slope of the trace of $f(x,y)$ for the plane $x = a$ at the point (a,b).
Example 2: Find the slopes of the traces to $z = 10 - 4x^2 - y^2$ at the point $(1,2)$
We sketch the traces for the planes $x = 1$ and $y = 2$:



Next, we'll need the two partial derivatives so we can get the slopes,

$$f_x(x,y) = -8x, \quad f_y(x,y) = -2y$$

To get the slopes, all we need to do is evaluate the partial derivatives at the point in question,

$$f_x(1,2) = -8, \quad f_y(1,2) = -4$$

So, the tangent line at $(1,2)$ for the trace to $z = 10 - 4x^2 - y^2$ for the planes $y = 2$ has a slope of -8. Also the tangent line at $(1,2)$ for the trace to $z = 10 - 4x^2 - y^2$ for the plane $x = 1$ has a slope of -4.
Finally, let's briefly talk about getting the equations of the tangent line. Recall that the **equation of a line** in 3-D space is given by a vector equation.
The point will be,

$$\big(a, b, f(a,b)\big)$$

The parallel (or tangent) vector is also just as easy,

$$\vec{r}(x,y) = \langle x, y, z\rangle = \langle x, y, f(x,y)\rangle$$

Here is the tangent vector for traces with fixed y,

$$\vec{r}_x(x,y) = \langle 1, 0, f_x(x,y)\rangle$$

For traces with fixed x the tangent vector is,

$$\vec{r}_y(x,y) = \langle 0,1,f_y(x,y)\rangle$$

The equation for the tangent line to traces with fixed y is then,

$$\vec{r}(t) = \langle a,b,f(a,b)\rangle + t\langle 1,0,f_x(a,b)\rangle$$

and the tangent line to traces with fixed x is,

$$\vec{r}(t) = \langle a,b,f(a,b)\rangle + t\langle 0,1,f_y(a,b)\rangle.$$

Higher Order Partial Derivatives

Just as we had higher order derivatives with functions of one variable, we will also have higher order derivatives of functions of more than one variable.

Consider the case of a function of two variables, $f(x,y)$ since both of the first order partial derivatives are also functions of x and y, we could in turn differentiate each with respect to x or y. This means that for the case of a function of two variables there will be a total of four possible second order derivatives.

$$(f_x)_x = f_{xx} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2}$$

$$(f_x)_y = f_{xy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x}$$

$$(f_y)_x = f_{yx} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y}$$

$$(f_y)_y = f_{yy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2}$$

The second and third second order partial derivatives are often called mixed partial derivatives since we are taking derivatives with respect to more than one variable. Note as well that the order that we take the derivatives in is given by the notation for each these. If we are using the subscripting notation, *e.g.* f_{xy}, then we will differentiate from left to right. In other words, in this case, we will differentiate first with respect to x and then with respect to y. With the fractional notation, *e.g.* $\frac{\partial^2 f}{\partial y \partial x}$, it is the opposite. In these cases, we differentiate moving along the denominator from right to left. So, again, in this case we differentiate with respect to x first and then y.

Clairaut's Theorem

> Suppose that f is defined on a disk D that contains the point (a,b). If the functions f_{xy} and f_{yx} are continuous on this disk then,
>
> $$f_{xy}(a,b) = f_{yx}(a,b)$$

So far, we have only looked at second order derivatives. There are, of course, higher order derivatives as well. Here are a couple of the third order partial derivatives of function of two variables.

$$f_{xyx} = (f_{xy})_x = \frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial y \partial x}\right) = \frac{\partial^3 f}{\partial x \partial y \partial x}$$

$$f_{yxx} = (f_{yx})_x = \frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial x \partial y}\right) = \frac{\partial^3 f}{\partial x^2 \partial y}$$

There is an extension to Clairaut's Theorem that says if all of these are continuous, then they should all be equal,

$$f_{xyx} = f_{yxx}$$

In general, we can extend Clairaut's theorem to any function and mixed partial derivatives.

Differentials

Given the function $z = f(x,y)$ the differential dz or df is given by,

$$dz = f_x dx + f_y dy \quad \text{or} \quad df = f_x dx + f_y dy$$

There is a natural extension to functions of three or more variables. For instance, given the function $w = g(x,y,z)$ the differential is given by,

$$dw = g_x dx + g_y dy + g_z dz$$

Note that sometimes these differentials are called the **total differentials**.

Chain Rule

Case 1: $z = f(x,y)$, $x = g(t)$, $y = h(t)$ and compute $\frac{dz}{dt}$.

In this case we are going to compute an ordinary derivative, since z really would be a function of t only if we were to substitute in for x and y.

The chain rule for this case is,

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

So, basically what we're doing here is differentiating f with respect to each variable in it, and then multiplying each of these by the derivative of that variable with respect to t. The final step is to add all this up.

Example 1: Compute $\frac{dz}{dt}$ for each of the following.

(a) $z = xe^{xy}$, $x = t^2$, $y = t^{-1}$

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

$$= 2t(e^{xy} + yxe^{xy}) - t^{-2}x^2e^{xy}$$

So, technically we've computed the derivative. However, we should probably go ahead and substitute in for x and y as well at this point, since we've already got t's in the derivative. Doing this gives,

$$\frac{dz}{dt} = 2t(e^t + te^t) - t^{-2}t^4e^t = 2te^t + t^2e^t$$

Note that in this case it might actually have been easier to just substitute in for x and y in the original function, and just compute the derivative as we normally would. For comparison's sake let's do that.

$$z = t^2e^t \quad \Longrightarrow \quad \frac{dz}{dt} = 2te^t + t^2e^t$$

The same result for less work. Note however, that often it will actually be more work to do the substitution first.

(b) $z = x^2y^3 + y\cos x$, $x = \ln(t^2)$, $y = \sin(4t)$

$$\frac{dz}{dt} = (2xy^3 - y\sin x)\left(\frac{2}{t}\right) + (3x^2y^2 + \cos x)(4\cos(4t))$$

Now, there is a special case that we should take a quick look at before moving on to the next case. Let's suppose that we have the following situation,

$$z = f(x,y) \quad y = g(x)$$

In this case the chain rule for $\frac{dz}{dx}$ becomes,

$$\frac{dz}{dx} = \frac{\partial f}{\partial x}\frac{dx}{dx} + \frac{\partial f}{\partial y}\frac{dy}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}$$

In the first term we are using the fact that,

$$\frac{dx}{dx} = \frac{d}{dx}(x) = 1$$

Case 2: $z = f(x, y)$, $x = g(s, t)$, $y = h(s, t)$ and compute $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$.

In this case if we were to substitute in for x and y we would get that z is a function of s and t, and so it makes sense that we would be computing partial derivatives here and that there would be two of them.

Here is the chain rule for both of these cases.

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} \quad \text{and} \quad \frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$$

Let's see the general version of the chain rule.

Chain Rule

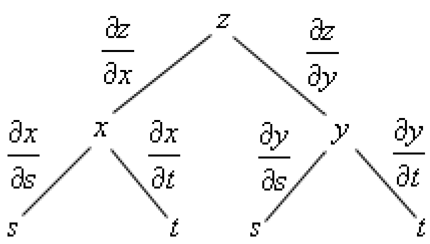
Suppose that z is a function of n variables, $x_1, x_2, \cdots, x_n$, and that each of these variables are in turn functions of m variables, $t_1, t_2, \cdots, t_m$. Then for any variable t_i, $i = 1,2,\cdots, m$, we have the following,

$$\frac{\partial z}{\partial t_i} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial z}{\partial x_2}\frac{\partial x_2}{\partial t_i} + \cdots + \frac{\partial z}{\partial x_n}\frac{\partial x_n}{\partial t_i}$$

Wow. That's a lot to remember. There is actually an easier way to construct all the chain rules that we've discussed in the section, or will look at in later examples. We can build up a **tree diagram** that will give us the chain rule for any situation. To see how these works, let's go back and take a look at the chain rule for $\frac{\partial z}{\partial s}$ given that $z = f(x, y)$, $x = g(s, t)$, $y = h(s, t)$. We already know what this is, but it may help to illustrate the tree diagram if we already know the answer. For reference here is the chain rule for this case,

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}$$

Here is the tree diagram for this case.



Note that the letter in the numerator of the partial derivative is the upper "node" of the tree, and the letter in the denominator of the partial derivative is the lower "node" of the tree.

We've now seen how to take first derivatives of these more complicated situations, but what about higher order derivatives? (it's same, just treat first order derivatives as functions)

The final topic in this section is a revisiting of **implicit differentiation**.

we will start off with a function in the form $F(x,y,z)=0$ and assume that $z=f(x,y)$ and let's start by trying to find $\frac{\partial z}{\partial x}$.

$$\frac{\partial F}{\partial x}\frac{\partial x}{\partial x}+\frac{\partial F}{\partial y}\frac{\partial y}{\partial x}+\frac{\partial F}{\partial z}\frac{\partial z}{\partial x}=0$$

Now, we have the following,

$$\frac{\partial x}{\partial x}=1 \quad \text{and} \quad \frac{\partial y}{\partial x}=0$$

Thus

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

Directional Derivatives

To this point we've only looked at the two partial derivatives $f_x(x,y)$ and $f_y(x,y)$.

Vectors can be used to define a direction and so the particle, at this point, can be said to be moving in the direction,

$$\vec{v}=\langle 2,1\rangle$$

In this way we will know that x is increasing twice as fast as y is.

We need a way to consistently find the rate of change of a function in a given direction. We will do this by insisting that, the vector that defines the direction of change be a unit vector.

$$\vec{v}=\langle\frac{2}{\sqrt{5}},\frac{1}{\sqrt{5}}\rangle$$

since this is the unit vector that points in the direction of change.

Sometimes we will give the direction of changing x and y as an angle. For instance, we may say that we want the rate of change of f in the direction of $\theta=\frac{\pi}{3}$. The unit vector that points in this direction is given by,

$$\vec{u}=\langle\cos\theta,\sin\theta\rangle$$

Okay, now that we know how to define the direction of changing x and y, it's time to start talking about finding the rate of change of f in this direction. Let's start off with the official definition.

The rate of change of $f(x,y)$ in the direction of the unit vector $\vec{u}=\langle a,b\rangle$ is called the **directional derivative** and is denoted by $D_{\vec{u}}f(x,y)$. The definition of the directional derivative is,

$$D_{\vec{u}}f(x,y)=\lim_{h\to 0}\frac{f(x+ah,y+bh)-f(x,y)}{h}$$

In practice this can be a very difficult limit to compute so we need an easier way of taking directional derivatives.

Let's define a new function of a single variable,

$$g(z)=f(x_0+az,y_0+bz)$$

where x_0, y_0, a, and b are some fixed numbers.

$$g'(z)=\lim_{h\to 0}\frac{g(z+h)-g(z)}{h}$$

$$g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h}$$
$$= \lim_{h \to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0)}{h}$$
$$= D_{\vec{u}} f(x_0, y_0)$$

So, it looks like we have the following relationship.

$$g'(0) = D_{\vec{u}} f(x_0, y_0)$$

Let's rewrite $g(z)$ as follows,

$$g(z) = f(x, y) \quad \text{where } x = x_0 + az \quad \text{and} \quad y = y_0 + bz$$
$$g'(z) = \frac{dg}{dz} = \frac{\partial f}{\partial x}\frac{dx}{dz} + \frac{\partial f}{\partial y}\frac{dy}{dz} = f_x(x, y)a + f_y(x, y)b$$
$$g'(0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b$$

Now,

$$D_{\vec{u}} f(x_0, y_0) = g'(0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b$$

If we go back to allowing x and y to be any number, we get the following formula for computing directional derivatives.

$$\boxed{D_{\vec{u}} f(x, y) = f_x(x, y)a + f_y(x, y)b}$$

This is much simpler than the limit definition. Also note that this definition assumed that we were working with functions of two variables. There are similar formulas that can be derived by the same type of argument for functions with more than two variables. For instance, the directional derivative of $f(x, y, z)$ in the direction of the unit vector $\vec{u} = \langle a, b, c \rangle$ is given by,

$$\boxed{D_{\vec{u}} f(x, y, z) = f_x(x, y, z)a + f_y(x, y, z)b + f_z(x, y, z)c}$$

There is another form of the formula that we used to get the directional derivative. That is a little nicer and somewhat more compact.

$$D_{\vec{u}} f(x, y, z) = f_x(x, y, z)a + f_y(x, y, z)b + f_z(x, y, z)c$$
$$= \langle f_x, f_y, f_z \rangle \cdot \langle a, b, c \rangle$$

In other words, we can write the directional derivative as a dot product and notice that the second vector is nothing more than the unit vector $\vec{u}$ that gives the direction of change.

Now let's give a name and notation to the first vector in the dot product, since this vector will show up fairly regularly throughout this course (and in other courses). The **gradient of f** or **gradient vector of f** is defined to be,

$$\nabla f = \langle f_x, f_y, f_z \rangle \quad \text{or} \quad \nabla f = \langle f_x, f_y \rangle$$

Or, if we want to use the standard basis vectors the gradient is,

$$\nabla f = f_x \vec{i} + f_y \vec{j} + f_z \vec{k} \quad \text{or} \quad \nabla f = f_x \vec{i} + f_y \vec{j}$$

there is a natural extension to functions of any number of variables that we'd like.

With the definition of the gradient we can now say that the directional derivative is given by,

$$D_{\vec{u}} f = \nabla f \cdot \vec{u}$$

where we will no longer show the variable and use this formula for any number of variables. Note as well that we will sometimes use the following notation,

$$D_{\vec{u}} f(\vec{x}) = \nabla f \cdot \vec{u}$$

where $\vec{x} = \langle x, y, z \rangle$ or $\vec{x} = \langle x, y \rangle$ as needed. $\vec{x}$ will be used to represent as many variables as we need.

Before proceeding, let's note that the first order partial derivatives that we were looking at in the majority of the section can be thought of as special cases of the directional derivatives. For instance, f_x can be thought of as the directional derivative of f in the direction of $\vec{u} = \langle 1,0 \rangle$ or $\vec{u} = \langle 1,0,0 \rangle$, depending on the number of variables that we're working with. The same can be done for f_y and f_z.

Theorem

> The maximum value of $D_{\vec{u}} f(\vec{x})$ (and hence then the maximum rate of change of the function $f(\vec{x})$) is given by $\|\nabla f(\vec{x})\|$ and will occur in the direction given by $\nabla f(\vec{x})$. ($D_{\vec{u}} f = \nabla f \cdot \vec{u} = \|\nabla f\| \|\vec{u}\| \cos\theta = \|\nabla f\| \cos\theta$)

Fact

> The gradient vector $\nabla f(x_0, y_0)$ is orthogonal (or perpendicular) to the **level curve** $f(x,y) = k$ at the point (x_0, y_0). Likewise, the gradient vector $\nabla f(x_0, y_0, z_0)$ is orthogonal to the **level surface** $f(x,y,z) = k$ at the point (x_0, y_0, z_0).

Proof

We're going to do the proof for the R^3 case. The proof for the R^2 case is identical. Let's S be the level surface given by $f(x,y,z) = k$ and let $P = (x_0, y_0, z_0)$. Note as well that P will be on S.

Now, let C be any curve on S that contains P. Let $\vec{r}(t) = \langle x(t), y(t), z(t) \rangle$ be the vector equation for C, and suppose that t_0 be the value of t such that $\vec{r}(t_0) = \langle x_0, y_0, z_0 \rangle$.

Because C lies on S, so

$$f\big(x(t), y(t), z(t)\big) = k$$

Next, let's use the Chain Rule on this to get,

$$\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = 0$$

Notice that $\nabla f = \langle f_x, f_y, f_z \rangle$ and $\vec{r}'(t) = \langle x'(t), y'(t), z'(t) \rangle$. So this becomes,

$$\nabla f \cdot \vec{r}'(t) = 0$$

At $t = t_0$ this is,

$$\nabla f(x_0, y_0, z_0) \cdot \vec{r}'(t_0) = 0$$

This then tells us that the gradient vector at $P = (x_0, y_0, z_0)$, $\nabla f(x_0, y_0, z_0)$ is orthogonal to the tangent vector, $\vec{r}'(t_0)$, to any curve C that passes through P and on the surface S, and so must also be orthogonal to the surface S.

Applications of Partial Derivatives

Relative Minimums and Maximums

We are going to be looking at identifying relative minimums and relative maximums. Recall that we will often use the word extrema to refer to both minimums and maximums.

Definition

> 1. A function $f(x,y)$ has a **relative minimum** at the point (a,b) if $f(x,y) \geq f(a,b)$ for all points (x,y) in some region around (a,b).
> 2. A function $f(x,y)$ has a **relative maximum** at the point (a,b) if $f(x,y) \leq f(a,b)$ for all points (x,y) in some region around (a,b).

Next, we need to extend the idea of **critical points** up to functions of two variables. Recall that a critical point of the function $f(x)$ was a number $x = c$ so that either $f'(c) = 0$ or $f'(c)$ doesn't exist. We have a similar definition for critical points of functions of two variables.

Definition

> The point (a, b) is a **critical point** (or a **stationary point**) of $f(x, y)$ provided one of the following is true,
> 1. $\nabla f(a, b) = \vec{0}$ (this is equivalent to saying that $f_x(a, b) = 0$ and $f_y(a, b) = 0$),
> 2. $f_x(a, b)$ and/or $f_y(a, b)$ doesn't exist.

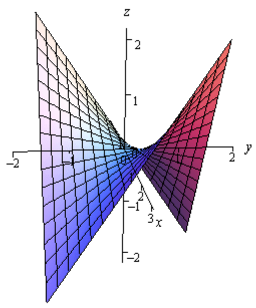
If only one of the first order partial derivatives are zero at the point then the point will NOT be a critical point.

Fact

> If the point (a, b) is a relative extrema of the function $f(x, y)$ and the first order derivatives of $f(x, y)$ exist at (a, b) then (a, b) is also a critical point of $f(x, y)$ and in fact we'll have $\nabla f(a, b) = \vec{0}$.

Note that this does NOT say that all critical points are relative extrema. It only says that relative extrema will be critical points of the function. To see this let's consider the function

$$f(x, y) = xy$$
$$f_x(x, y) = y \quad \text{and} \quad f_y(x, y) = y$$



there is no way that $(0,0)$ can be a relative extrema.
Critical points that exhibit this kind of behavior are called **saddle points**.
So, once we have all the critical points in hand, all we will need to do is test these points to see if they are relative extrema or not.

Fact

> Suppose that (a, b) is a critical point of $f(x, y)$ and that the second order partial derivatives are continuous in some region that contains (a, b). Next define, (***Hessian***)
>
> $$D = D(a, b) = f_{xx}(a, b) f_{yy}(a, b) - \left[f_{xy}(a, b)\right]^2$$
>
> We then have the following **classifications** of the critical point.
> 1. If $D > 0$ and $f_{xx}(a, b) > 0$ then there is a relative minimum at (a, b).
> 2. If $D > 0$ and $f_{xx}(a, b) < 0$ then there is a relative maximum at (a, b).
> 3. If $D < 0$ then the point (a, b) is a saddle point.
> 4. If $D = 0$ then the point (a, b) may be a relative minimum, relative maximum or a saddle point. Other techniques would need to be used to classify the critical point.

Note that if $D > 0$ then both $f_{xx}(a,b)$ and $f_{yy}(a,b)$ will have the same sign, and so in the first two cases above, we could just as easily replace $f_{xx}(a,b)$ with $f_{yy}(a,b)$. Also note that we aren't going to be seeing any cases in this class where $D = 0$, as these can often be quite difficult to classify.

Example 2: Find and classify all the critical points for $f(x,y) = 3x^2y + y^3 - 3x^2 - 3y^2 + 2$ we will first need to get all the first and second order derivatives.

$$f_x = 6xy - 6x \quad f_y = 3x^2 + 3y^2 - 6y$$

$$f_{xx} = 6y - 6 \quad f_{yy} = 6y - 6 \quad f_{xy} = 6x$$

We'll first need the critical points. The equations that we'll need to solve this time are,

$$6xy - 6x = 0$$

$$3x^2 + 3y^2 - 6y = 0$$

First, let's notice that we can factor out a $6x$ from the first equation to get,

$$6x(y-1) = 0$$

So, we can see that the first equation will be zero if $x = 0$ or $y = 1$. Be careful to not just cancel the x from both sides. If we had done that, we would have missed $x = 0$.

To find the critical points we can plug these (individually) into the second equation and solve for the remaining variable.

$x = 0$:

$$3y^2 - 6y = 3y(y-2) = 0 \quad \Rightarrow \quad y = 0, y = 2$$

$y = 1$:

$$3x^2 - 3 = 3(x^2 - 1) = 0 \quad \Rightarrow \quad x = -1, x = 1$$

So, if $x = 0$ we have the following critical points,

$$(0,0) \quad (0,2)$$

and if $y = 1$ the critical points are,

$$(1,1) \quad (-1,1)$$

Now all we need to do is classify the critical points. To do this we'll need the general formula for D.

$$D(x,y) = (6y-6)(6y-6) - (6x)^2 = (6y-6)^2 - 36x^2$$

$(0,0)$:

$$D = D(0,0) = 36 > 0 \quad f_{xx}(0,0) = -6 < 0$$

$(0,2)$:

$$D = D(0,2) = 36 > 0 \quad f_{xx}(0,0) = 6 > 0$$

$(1,1)$:

$$D = D(1,1) = -36 < 0$$

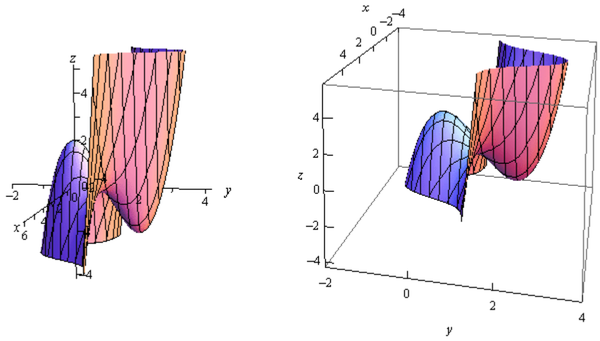
$(-1,1)$:

$$D = D(-1,1) = -36 < 0$$

So, it looks like we have the following classification of each of these critical points.

$$\begin{array}{ccc} (0,0) & : & \text{Relative Maximum} \\ (0,2) & : & \text{Relative Minimum} \\ (1,1) & : & \text{Saddle Point} \\ (-1,1) & : & \text{Saddle Point} \end{array}$$

Here is a graph of the surface for the sake of completeness.



Example 3: Determine the point on the plane $4x-2y+z=1$ that is closest to the point $(-2,-1,5)$. (answer: $\left(-\frac{34}{21},-\frac{25}{21},\frac{107}{21}\right)$)

Absolute Extrema

In this section we want to optimize a function, that is identify the absolute minimum and/or the absolute maximum of the function, on a given region in R^2. Note that when we say we are going to be working on a region in R^2 we mean that we're going to be looking at some region in the xy-plane.

In order to optimize a function in a region, we are going to need to get a couple of definitions out of the way and a fact.

Definitions

1. A region in R^2 is called **closed** if it includes its boundary. A region is called **open** if it doesn't include any of its boundary points.
2. A region in R^2 is called **bounded** if it can be completely contained in a disk. In other words, a region will be bounded if it is finite.

Let's think a little more about the definition of closed. Below are two definitions of a rectangle, one is closed and the other is open.

$$\begin{array}{cc} \text{Open} & \text{Closed} \\ -5<x<3 & -5\le x\le 3 \\ 1<y<6 & 1\le y\le 6 \end{array}$$

In this first case we don't allow the ranges to include the endpoints (*i.e.* we aren't including the edges of the rectangle), and so we aren't allowing the region to include any points on the edge of the rectangle. In other words, we aren't allowing the region to include its boundary and so it's open.

In the second case we are allowing the region to contain points on the edges and so will contain its entire boundary, and hence will be closed.

Extreme Value Theorem

If $f(x,y)$ is continuous in some closed, bounded set D in R^2 then there are points in D, (x_1,y_1) and (x_2,y_2) so that $f(x_1,y_1)$ is the absolute maximum and $f(x_2,y_2)$ is the absolute minimum of the function in D.

Note that this theorem does NOT tell us where the absolute minimum or absolute maximum will occur.

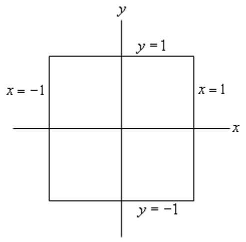
Note as well that the absolute minimum and/or absolute maximum may occur in the interior of the region or it may occur on the boundary of the region.

Finding Absolute Extrema

1. Find all the critical points of the function that lie in the region D and determine the function value at each of these points.
2. Find all extrema of the function on the boundary. This usually involves the Calculus I approach for this work.
3. The largest and smallest values found in the first two steps are the absolute minimum and the absolute maximum of the function.

Example 1: Find the absolute minimum and absolute maximum of $f(x,y) = x^2 + 4y^2 - 2x^2y + 4$ on the rectangle given by $-1 \le x \le 1$ and $-1 \le y \le 1$.

Let's first get a quick picture of the rectangle for reference purposes.



The boundary of this rectangle is given by the following conditions.

$$\begin{aligned} \text{right side:} &\quad x = 1, -1 \le y \le 1 \\ \text{left side:} &\quad x = -1, -1 \le y \le 1 \\ \text{upper side:} &\quad y = 1, -1 \le x \le 1 \\ \text{lower side:} &\quad y = -1, -1 \le x \le 1 \end{aligned}$$

We'll start this off by finding all the critical points that lie inside the given rectangle. To do this we'll need the two first order derivatives.

$$f_x = 2x - 4xy \quad \text{and} \quad f_y = 8y - 2x^2$$

Note that since we aren't going to be classifying the critical points, we don't need the second order derivatives. To find the critical points we will need to solve the system,

$$2x - 4xy = 0$$

$$8y - 2x^2 = 0$$

We can get

$$y = \frac{x^2}{4}$$

$$2x - 4x\left(\frac{x^2}{4}\right) = 2x - x^3 = x(2 - x^2) = 0$$

This tells us that we must have $x = 0$ or $x = \pm\sqrt{2} = \pm 1.414\cdots$. Now, recall that we only want critical points in the region that we're given. That means that we only want critical points for which $-1 \le x \le 1$. Note however that a simple change to the boundary would include these two, so don't forget to always check if the critical points are in the region (or on the boundary since that can also happen).

Plugging $x = 0$ into the equation for y gives us,

$$y = \frac{0^2}{4} = 0$$

The single critical point, in the region (and again, that's important), is $(0,0)$. We now need to get the value of the function at the critical point.

$$f(0,0) = 4$$

Now we have reached the long part of this problem. We need to find the absolute extrema of the function along the boundary of the rectangle.

Let's first take a look at the right side. As noted above the right side is defined by

$$x = 1, -1 \le y \le 1$$

Let's take advantage of $x = 1$ by defining a new function as follows,

$$g(y) = f(1, y) = 1^2 + 4y^2 - 2(1^2)y + 4 = 5 + 4y^2 - 2y$$

Now, finding the absolute extrema of $f(x, y)$ along the right side will be equivalent to finding the absolute extrema of $g(y)$ in the range $-1 \le y \le 1$. We find the critical points of $g(y)$ in the range $-1 \le y \le 1$, and then evaluate $g(y)$ at the critical points and the end points of the range of y's.

Let's do that for this problem.

$$g'(y) = 8y - 2 \quad \Rightarrow \quad y = \frac{1}{4}$$

This is in the range and so we will need the following function evaluations.

$$g(-1) = 11 \quad g(1) = 7 \quad g\left(\frac{1}{4}\right) = \frac{19}{4} = 4.75$$

Notice that, using the definition of $g(y)$, these are also function values for $f(x, y)$.

$$g(-1) = f(1, -1) = 11$$
$$g(1) = f(1,1) = 7$$
$$g\left(\frac{1}{4}\right) = f\left(1, \frac{1}{4}\right) = \frac{19}{4} = 4.75$$

We can now do the left side of the rectangle which is defined by,

$$x = -1, -1 \le y \le 1$$
$$g(y) = f(-1, y) = (-1)^2 + 4y^2 - 2(-1)^2y + 4 = 5 + 4y^2 - 2y$$

Notice however that, for this boundary, this is the same function as we looked at for the right side. This will not always happen, but since it has, let's take advantage of the fact that we've already done the work for this function.

$$g(-1) = 11 \quad g(1) = 7 \quad g\left(\frac{1}{4}\right) = \frac{19}{4} = 4.75$$

The only real difference here is that these will correspond to values of $f(x, y)$ at different points than for the right side.

$$g(-1) = f(-1, -1) = 11$$
$$g(1) = f(-1,1) = 7$$
$$g\left(\frac{1}{4}\right) = f\left(-1, \frac{1}{4}\right) = \frac{19}{4} = 4.75$$

We can now look at the upper side defined by,

$$y = 1, -1 \le x \le 1$$
$$h(x) = f(x, 1) = x^2 + 4(1^2) - 2x^2(1) + 4 = 8 - x^2$$

First find the critical point(s).

$$h'(x) = -2x \quad \Rightarrow \quad x = 0$$
$$h(-1) = 7 \quad h(1) = 7 \quad h(0) = 8$$

and these in turn correspond to the following function values for $f(x, y)$

$$h(-1) = f(-1,1) = 7$$
$$h(1) = f(1,1) = 7$$
$$h(0) = f(0,1) = 8$$

Note that there are several "repeats" here. The first two function values have already been computed when we looked at the right and left side. This will often happen.

Finally, we need to take care of the lower side. This side is defined by,

$$y = -1, -1 \leq x \leq 1$$
$$h(x) = f(x, -1) = x^2 + 4(-1)^2 - 2x^2(-1) + 4 = 8 + 3x^2$$
$$h'(x) = 6x \quad \Rightarrow \quad x = 0$$
$$h(-1) = 11 \quad h(1) = 11 \quad h(0) = 8$$

and the corresponding values for $f(x, y)$ are,

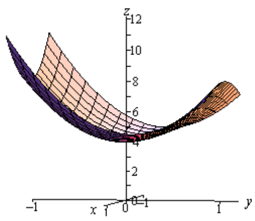
$$h(-1) = f(-1, -1) = 11$$
$$h(1) = f(1, -1) = 11$$
$$h(0) = f(0, -1) = 8$$

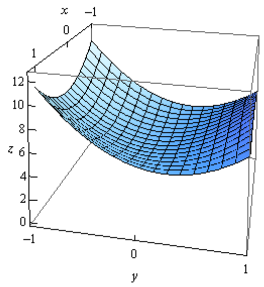
The final step to this (long…) process is to collect up all the function values for $f(x, y)$ that we've computed in this problem. Here they are,

$$\begin{array}{ccc} f(0,0) = 4 & f(1,-1) = 11 & f(1,1) = 7 \\ f(1,1/4) = 4.75 & f(-1,1) = 7 & f(-1,-1) = 11 \\ f(-1,1/4) = 4.75 & f(0,1) = 8 & f(0,-1) = 8 \end{array}$$

The absolute minimum is at $(0,0)$ since gives the smallest function value and the absolute maximum occurs at $(1, -1)$ and $(-1, -1)$ since these two points give the largest function value.

Here is a sketch of the function on the rectangle for reference purposes.





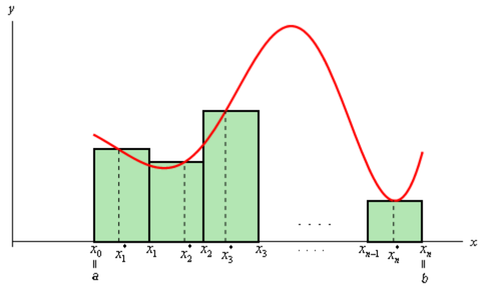
Double Integrals

Before starting on double integrals, let's do a quick review of the definition of definite integrals for functions of single variables. First, when working with the integral,

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

we think of x's as coming from the interval $a \le x \le b$. For these integrals we can say that we are integrating over the interval $a \le x \le b$.

Now, when we derived the definition of the definite integral, we first thought of this as an area problem. We first asked what the area under the curve was and to do this we broke up the interval $a \le x \le b$ into n subintervals of width Δx and choose a point, x_i^*, from each interval as shown below,



Each of the rectangles has height of $f(x_i^*)$ and we could then use the area of each of these rectangles to approximate the area as follows.

$$A \approx f(x_1^*)\Delta x + f(x_2^*)\Delta x + \cdots + f(x_i^*)\Delta x + \cdots + f(x_n^*)\Delta x$$

To get the exact area, we then took the limit as n goes to infinity and this was also the definition of the definite integral.

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

In this section we want to integrate a function of two variables, $f(x, y)$. With functions of one variable we integrated over an interval (*i.e.* a one-dimensional space), and so it makes some sense that when integrating a function of two variables, we will integrate over a region of R^2 (two-dimensional space).

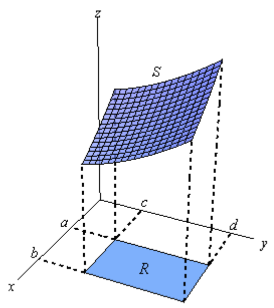
We will start out by assuming that the region in R^2 is a rectangle, which we will denote as follows,

$$R = [a, b] \times [c, d]$$

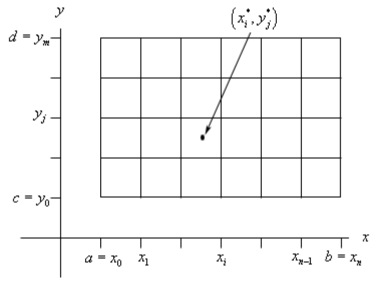
This means that the ranges for x and y are $a \le x \le b$ and $c \le y \le d$.

Also, we will initially assume that $f(x, y) \ge 0$, although this doesn't really have to be the case.

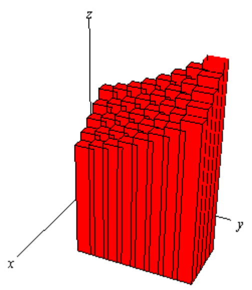
Let's start out with the graph of the surface S given by graphing $f(x, y)$ over the rectangle R.



Let's first ask what the volume of the region under S (and above the xy-plane of course) is. We will approximate the volume much as we approximated the area above. We will first divide up $a \le x \le b$ into n subintervals and divide up $c \le y \le d$ into m subintervals. This will divide up R into a series of smaller rectangles, and from each of these we will choose a point (x_i^*, y_j^*). Here is a sketch of this set up.



Now, over each of these smaller rectangles we will construct a box whose height is given by $f(x_i^*, y_j^*)$. Here is a sketch of that.



Each of the rectangles has a base area of ΔA and a height of $f(x_i^*, y_j^*)$, so the volume of each of these boxes is $f(x_i^*, y_j^*)\Delta A$. The volume under the surface S is then approximately,

$$V \approx \sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i^*, y_j^*)\Delta A$$

We will have a *double* sum since we will need to add up volumes in both the x and y directions.

To get a better estimation of the volume, we will take n and m larger and larger, and to get the exact volume we will need to take the limit as both n and m go to infinity. In other words,

$$V = \lim_{n,m \to \infty} \sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i^*, y_j^*)\Delta A$$

Here is the official definition of a double integral of a function of two variables over a rectangular region R, as well as the notation that we'll use for it.

$$\iint_R f(x,y)\,dA = \lim_{n,m\to\infty} \sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i^*, y_j^*)\,\Delta A$$

Note the similarities and differences in the notation to single integrals. Note that the differential is dA. Note as well that we don't have limits on the integrals in this notation. Instead we have the R written below the two integrals to denote the region that we are integrating over.

As indicated above, one interpretation of the double integral of $f(x,y)$ over the rectangle R is the volume under the function $f(x,y)$ (and above the xy-plane). Or,

$$\boxed{\text{Volume} = \iint_R f(x,y)\,dA}$$

Iterated Integrals

The following theorem tells us how to compute a double integral over a rectangle.

Fubini's Theorem

If $f(x,y)$ is continuous on $R = [a,b] \times [c,d]$ then,

$$\iint_R f(x,y)\,dA = \int_a^b \int_c^d f(x,y)\,dy\,dx = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

These integrals are called **iterated integrals**.

Note that there are in fact two ways of computing a double integral over a rectangle, and also notice that the inner differential matches up with the limits on the inner integral and similarly for the outer differential and limits.

Now, on some level this is just notation and doesn't really tell us how to compute the double integral. Let's just take the first possibility above and change the notation a little.

$$\iint_R f(x,y)\,dA = \int_a^b \left[\int_c^d f(x,y)\,dy\right] dx$$

We will compute the double integral by first computing

$$\int_c^d f(x,y)\,dy$$

and we compute this by holding x constant and integrating with respect to y, as if this were a single integral. This will give a function involving only x's which we can in turn integrate.

We've done a similar process with partial derivatives.

We can do the integral in either direction. However, sometimes one direction of integration is significantly **easier** than the other, so make sure that you think about which one you should do first before actually doing the integral.

There is a nice special case of this kind of integral. First, let's assume that $f(x,y) = g(x)h(y)$. Then, the integral becomes,

$$\iint_R f(x,y)\,dA = \iint_R g(x)h(y)\,dA = \int_c^d \int_a^b g(x)h(y)\,dx\,dy$$

Note that it doesn't matter in this case which variable we integrate first, as either order will arrive at the same result with the same work.
Next, notice

$$\iint_R f(x,y)dA = \iint_R g(x)h(y)dA = \int_a^b g(x)dx \int_c^d h(y)dy$$

In other words, if we can break up the function into a function only of x times a function of only y, then we can do the two integrals individually and multiply them together.

Fact

If $f(x,y) = g(x)h(y)$ and we are integrating over the rectangle $R = [a,b] \times [c,d]$ then,

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

We have one more topic to discuss in this section.
What we want to do is discuss single indefinite integrals of a function of two variables. In other words, we want to look at integrals like the following.

$$\int (x\sec^2(2y) + 4xy)dy$$

$$\int \left(x^3 - e^{-\frac{x}{y}}\right)dx$$

In the case of the first integral, we are asking what function we differentiated with respect to y to get the integrand, while in the second integral we're asking what function differentiated with respect to x to get the integrand.
Here are the integrals.

$$\int (x\sec^2(2y) + 4xy)dy = \frac{x}{2}\tan(2y) + 2xy^2 + g(x)$$

$$\int \left(x^3 - e^{-\frac{x}{y}}\right)dx = \frac{1}{4}x^4 + ye^{-\frac{x}{y}} + h(y)$$

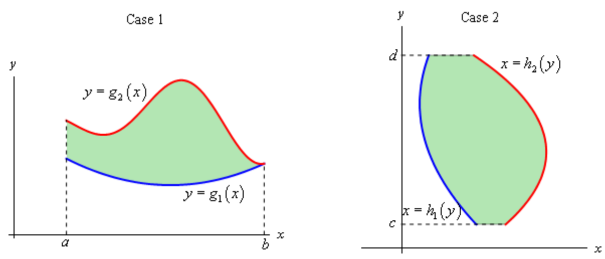
Notice that the "constants" of integration are now functions of the opposite variable.

Double Integrals Over General Regions

In the previous section we looked at double integrals over rectangular regions. The problem with this is that most of the regions are not rectangular, so we need to now look at the following double integral,

$$\iint_D f(x,y)dA$$

where D is any region.
There are two types of regions that we need to look at. Here is a sketch of both of them.



We will often use set *builder notation* to describe these regions. Here is the definition for the region in Case 1

$$D = \{(x,y)|a \le x \le b, g_1(x) \le y \le g_2(x)\}$$

and here is the definition for the region in Case 2.

$$D = \{(x,y)|h_1(y) \le x \le h_2(y), c \le y \le d\}$$

This notation is really just a fancy way of saying we are going to use all the points, (x,y), in which both of the coordinates satisfy the two given inequalities.

The double integral for both of these cases are defined in terms of iterated integrals as follows.

In Case 1 where $D = \{(x,y)|a \le x \le b, g_1(x) \le y \le g_2(x)\}$, the integral is defined to be,

$$\iint_D f(x,y)dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)dydx$$

In Case 2 where $D = \{(x,y)|h_1(y) \le x \le h_2(y), c \le y \le d\}$ the integral is defined to be,

$$\iint_D f(x,y)dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x,y)dxdy$$

Properties

1. $\iint_D f(x,y) + g(x,y)dA = \iint_D f(x,y)dA + \iint_D g(x,y)dA$

2. $\iint_D cf(x,y)dA = c\iint_D f(x,y)dA$, where c is any constant.

3. If the region D can be split into two separate regions D_1 and D_2 then the integral can be written as

$$\iint_D f(x,y)dA = \iint_{D_1} f(x,y)dA + \iint_{D_2} f(x,y)dA$$

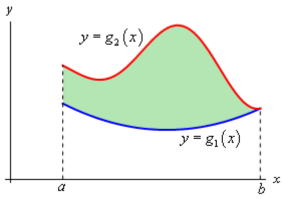
The first interpretation of a double integral is the volume of the solid, which lies below the surface given by $z = f(x,y)$ and above the region D in the xy-plane, is given by,

$$V = \iint_D f(x,y)dA$$

The second geometric interpretation of a double integral is the following.

$$\text{Area of } D = \iint_D dA$$

This is easy to see why this is true in general. Let's suppose that we want to find the area of the region shown below.



Double Integrals in Polar Coordinates

In this section we want to look at some regions that are much easier to describe in terms of polar coordinates.

$$\iint_D f(x,y)dA, \quad D \text{ is the disk of radius } 2$$

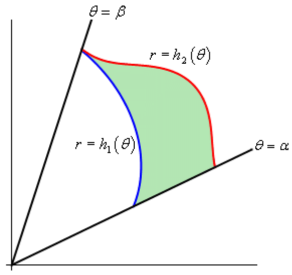
A disk of radius 2 can be defined in polar coordinates by the following inequalities,

$$0 \le \theta \le 2\pi$$

$$0 \le r \le 2$$

So, if we could convert our double integral formula into one involving polar coordinates, we would be in pretty good shape. The problem is that we can't just convert the dx and the dy into a dr and a $d\theta$. In computing double integrals to this point, we have been using the fact that $dA = dxdy$, and this really does require Cartesian coordinates to use. Once we've moved into polar coordinates $dA \neq drd\theta$, and so we're going to need to determine just what dA is under polar coordinates.

So, let's step back a little bit and start off with a general region in terms of polar coordinates, and see what we can do with that. Here is a sketch of some region using polar coordinates.

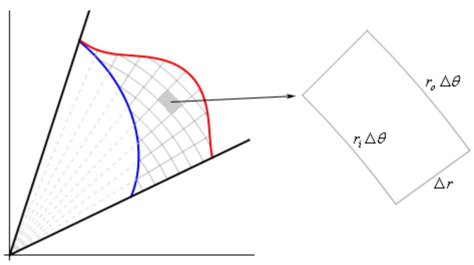


So, our general region will be defined by inequalities,

$$\alpha \le \theta \le \beta$$

$$h_1(\theta) \le r \le h_2(\theta)$$

Now, to find dA, let's redo the figure above as follows,



The area of this piece is ΔA. The two sides of this piece both have length $\Delta r = r_o - r_i$ where r_o is the radius of the outer arc and r_i is the radius of the inner arc. Basic geometry then tells us that the length of the inner edge is $r_i \Delta\theta$, while the length of the outer edge is $r_o \Delta\theta$ where $\Delta\theta$ is the angle between the two radial lines that form the sides of this piece.

Now, let's assume that we've taken the mesh so small that we can assume that $r_o \approx r_i = r$, and with this assumption, we can also assume that our piece is close enough to a rectangle that we can also assume that,

$$\Delta A \approx r \Delta\theta \Delta r$$

Also, if we assume that the mesh is small enough then we can also assume that,

$$dA \approx \Delta A \quad d\theta \approx \Delta\theta \quad dr \approx \Delta r$$

With these assumptions we then get $dA = r dr d\theta$.

As the mesh size gets smaller and smaller, the formula above becomes more and more accurate and so we can say that,

$$\boxed{dA = r dr d\theta}$$

We'll see another way of deriving this once we reach the Change of Variables section later in this chapter. This second way will not involve any assumptions either, and so it maybe a little better way of deriving this.

Now, if we're going to be converting an integral in Cartesian coordinates into an integral in polar coordinates, we are going to have to make sure that we've also converted all the x's and y's into polar coordinates as well. To do this we'll need to remember the following conversion formulas,

$$x = r\cos\theta \quad y = r\sin\theta \quad r^2 = x^2 + y^2$$

We are now ready to write down a formula for the double integral in terms of polar coordinates.

$$\boxed{\iint\limits_D f(x,y)\,dA = \int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} f(r\cos\theta, r\sin\theta)\, r\,dr\,d\theta}$$

Triple Integrals

We'll use a **triple integral** to integrate over a three-dimensional region. The notation for the general triple integrals is,

$$\iiint\limits_E f(x,y,z)\,dV$$

Let's start simple by integrating over the box,

$$B = [a,b] \times [c,d] \times [r,s]$$

The triple integral in this case is,

$$\iiint\limits_B f(x,y,z)\,dV = \int_r^s \int_c^d \int_a^b f(x,y,z)\,dx\,dy\,dz$$

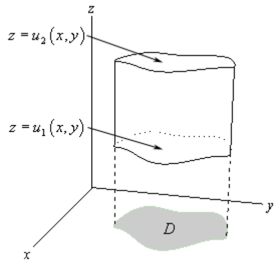
Before moving on to more general regions let's get a nice geometric interpretation about the triple integral out of the way, so we can use it in some of the examples to follow.

Fact

The volume of the three-dimensional region E is given by the integral,

$$V = \iiint_E dV$$

Let's now move on the more general three-dimensional regions. We have three different possibilities for a general region. Here is a sketch of the first possibility.



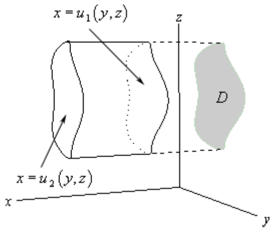
In this case we define the region E as follows,

$$E = \{(x,y,z) | (x,y) \in D, u_1(x,y) \le z \le u_2(x,y)\}$$

where $(x,y) \in D$ is the notation that means that the point (x,y) lies in the region D from the xy-plane. In this case we will evaluate the triple integral as follows,

$$\iiint_E f(x,y,z)dV = \iint_D \left[\int_{u_1(x,y)}^{u_2(x,y)} f(x,y,z)dz\right] dA$$

Let's now move onto the second possible three-dimensional region. Here is a sketch of this region.



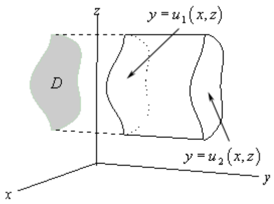
For this possibility we define the region E as follows,

$$E = \{(x,y,z) | (y,z) \in D, u_1(y,z) \le x \le u_2(y,z)\}$$

So, the region D will be a region in the yz-plane. Here is how we will evaluate these integrals.

$$\iiint_E f(x,y,z)dV = \iint_D \left[\int_{u_1(y,z)}^{u_2(y,z)} f(x,y,z)dx\right] dA$$

We now need to look at the third (and final) possible three-dimensional region. Here is a sketch of this region.



In this final case E is defined as,

$$E = \{(x,y,z) | (x,z) \in D, u_1(x,z) \le y \le u_2(x,z)\}$$

and here the region D will be a region in the xz-plane. Here is how we will evaluate these integrals.

$$\iiint_E f(x,y,z)dV = \iint_D \left[\int_{u_1(x,z)}^{u_2(x,z)} f(x,y,z)dy\right] dA$$

Triple Integrals in Cylindrical Coordinates

The following are the conversion formulas for cylindrical coordinates.

$$x = r\cos\theta \quad y = r\sin\theta \quad z = z$$

In order to do the integral in cylindrical coordinates, we will need to know what dV will become in terms of cylindrical coordinates. We will be able to show in the Change of Variables section of this chapter that,

$$\boxed{dV = r dz dr d\theta}$$

The region, E, over which we are integrating becomes,

$$E = \{(x,y,z)|(x,y) \in D, u_1(x,y) \le z \le u_2(x,y)\}$$
$$= \{(r,\theta,z)|\alpha \le \theta \le \beta, h_1(\theta) \le r \le h_2(\theta), u_1(r\cos\theta, r\sin\theta) \le z \le u_2(r\cos\theta, r\sin\theta)\}$$

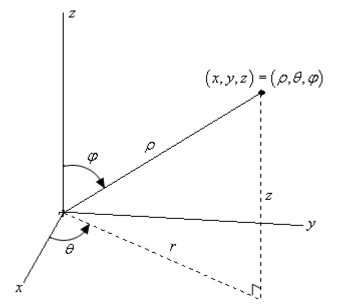
Note that we've only given this for E's in which D is in the xy-plane. We can modify this accordingly if D is in the yz-plane or the xz-plane as needed.

In terms of cylindrical coordinates, a triple integral is,

$$\boxed{\iiint_E f(x,y,z)dV = \int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} \int_{u_1(r\cos\theta, r\sin\theta)}^{u_2(r\cos\theta, r\sin\theta)} r\, f(r\cos\theta, r\sin\theta, z) dz dr d\theta}$$

Triple Integrals in Spherical Coordinates

First, we need to recall just how spherical coordinates are defined. The following sketch shows the relationship between the Cartesian and spherical coordinate systems.



Here are the conversion formulas for spherical coordinates.

$$x = \rho\sin\varphi\cos\theta, \quad y = \rho\sin\varphi\sin\theta, \quad z = \rho\cos\varphi$$
$$x^2 + y^2 + z^2 = \rho^2$$

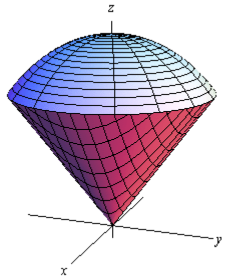
We also have the following restrictions on the coordinates.

$$\rho \ge 0, \quad 0 \le \varphi \le \pi$$

For our integrals we are going to restrict E down to a spherical wedge. This will mean that we are going to take ranges for the variables as follows,

$$a \le \rho \le b$$
$$\alpha \le \theta \le \beta$$
$$\delta \le \varphi \le \gamma$$

Most of the wedges we'll be working with will fit into this pattern.



In the next section we will show that

$$\boxed{dV = \rho^2 \sin\varphi \, d\rho d\theta d\varphi}$$

Therefore, the integral will become,

$$\boxed{\iiint\limits_E f(x,y,z)dV = \int_\delta^\gamma \int_\alpha^\beta \int_a^b \rho^2 \sin\varphi \, f(\rho\sin\varphi\cos\theta\,,\rho\sin\varphi\sin\theta\,,\rho\cos\varphi)\, d\rho d\theta d\varphi}$$

Change of Variables

While often the reason for changing variables is to get us an integral that we can do with the new variables, another reason for changing variables is to convert the region into a nicer region to work with.

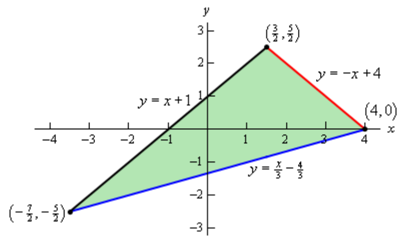
Before we move into changing variables with multiple integrals, we first need to see how the region may change with a change of variables.

First, we need a little terminology/notation out of the way. We call the equations that define the change of variables a **transformation**. Also, we will typically start out with a region, R, in xy-coordinates and transform it into a region in uv-coordinates.

Example 1: Determine the new region that we get by applying the given transformation to the region R.

(a) R is the region bounded by $y = -x + 4$, $y = x + 1$, and $y = \frac{x}{3} - \frac{4}{3}$ and the transformation is $x = \frac{1}{2}(u + v)$, $y = \frac{1}{2}(u - v)$.

Let's sketch the graph of the region and see what we've got,



Now, let's go through the transformation. Let's do $y = -x + 4$ first. Plugging in the transformation gives,

$$\frac{1}{2}(u - v) = -\frac{1}{2}(u + v) + 4$$

$$u - v = -u - v + 8$$

$$2u = 8$$

$$u = 4$$

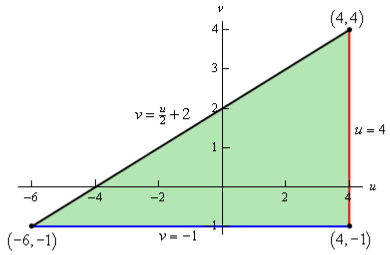
Now let's take a look at $y = x + 1$,

$$\frac{1}{2}(u - v) = \frac{1}{2}(u + v) + 1$$
$$u - v = u + v + 2$$
$$-2v = 2$$
$$v = -1$$

Finally, let's transform $y = \frac{x}{3} - \frac{4}{3}$,

$$\frac{1}{2}(u - v) = \frac{1}{3}\left(\frac{1}{2}(u + v)\right) - \frac{4}{3}$$
$$3u - 3v = u + v - 8$$
$$4v = 2u + 8$$
$$v = \frac{u}{2} + 2$$

Let's take a look at the new region that we get under the transformation.



We still get a triangle, but a much nicer one.

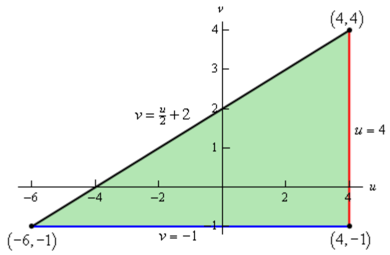
Before proceeding with the next topic, let's address another point. On occasion, we will also need to know the range of u and/or v for each of the new equations we get from the transformation.

So, let's work a quick example to see how we do that.

Example 2: For the region bounded by $y = -x + 4$, $y = x + 1$, and $y = \frac{x}{3} - \frac{4}{3}$ and the transformation $x = \frac{1}{2}(u + v)$, $y = \frac{1}{2}(u - v)$ determine the ranges of u and v for each of the new equations from the transformation.

Okay, we already know what the new region looks like and what the new equations are from the previous example.

Here is the new region we get under the transformation.



Let's now actually start working the problem.

Let's start with the equation $u = 4$ from $y = -x + 4$.

$$x = \frac{1}{2}(u + v) = x = \frac{1}{2}(4 + v)$$

Since $\frac{3}{2} \leq x \leq 4$ for $y = -x + 4$,

$$\frac{3}{2} \leq x \leq 4$$

$$\frac{3}{2} \leq \frac{1}{2}(4 + v) \leq 4$$

$$3 \leq 4 + v \leq 8$$

$$-1 \leq v \leq 4$$

Let's now move onto $v = -1$ from $y = x + 1$.

$$-\frac{7}{2} \leq x \leq \frac{3}{2}$$

$$-\frac{7}{2} \leq \frac{1}{2}(u - 1) \leq \frac{3}{2}$$

$$-7 \leq u - 1 \leq 3$$

$$-6 \leq u \leq 4$$

Finally, let's find the range of u and v for $v = \frac{u}{2} + 2$ from $y = \frac{x}{3} - \frac{4}{3}$.

$$-\frac{5}{2} \leq y \leq 0$$

$$-\frac{5}{2} \leq \frac{1}{2}\left(u - \left(\frac{u}{2} + 2\right)\right) \leq 0$$

$$-5 \leq \frac{u}{2} - 2 \leq 0$$

$$-3 \leq \frac{u}{2} \leq 2$$

$$-6 \leq u \leq 4$$

the appropriate range of v is as follows,

$$-6 \leq u \leq 4$$

$$-3 \leq \frac{u}{2} \leq 2$$

$$-1 \leq \frac{u}{2} + 2 \leq 4$$

$$-1 \leq v \leq 4$$

We need to now talk about how we actually do change of variables in the integral. We will start with double integrals. In order to change variables in a double integral, we will need the **Jacobian** of the transformation. Here is the definition of the Jacobian.

The **Jacobian** of the transformation $x = g(u, v)$, $y = h(u, v)$ is

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \det \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$

The Jacobian is defined as a **determinant** of a 2×2 matrix.
Another formula for the determinant is,

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

Proof: (From Khan academy)
First, we need to recall the concept of linear transformations in linear algebra (Chapter 4):

A mapping L from a vector space V into a vector space W is said to be a **linear transformation** if

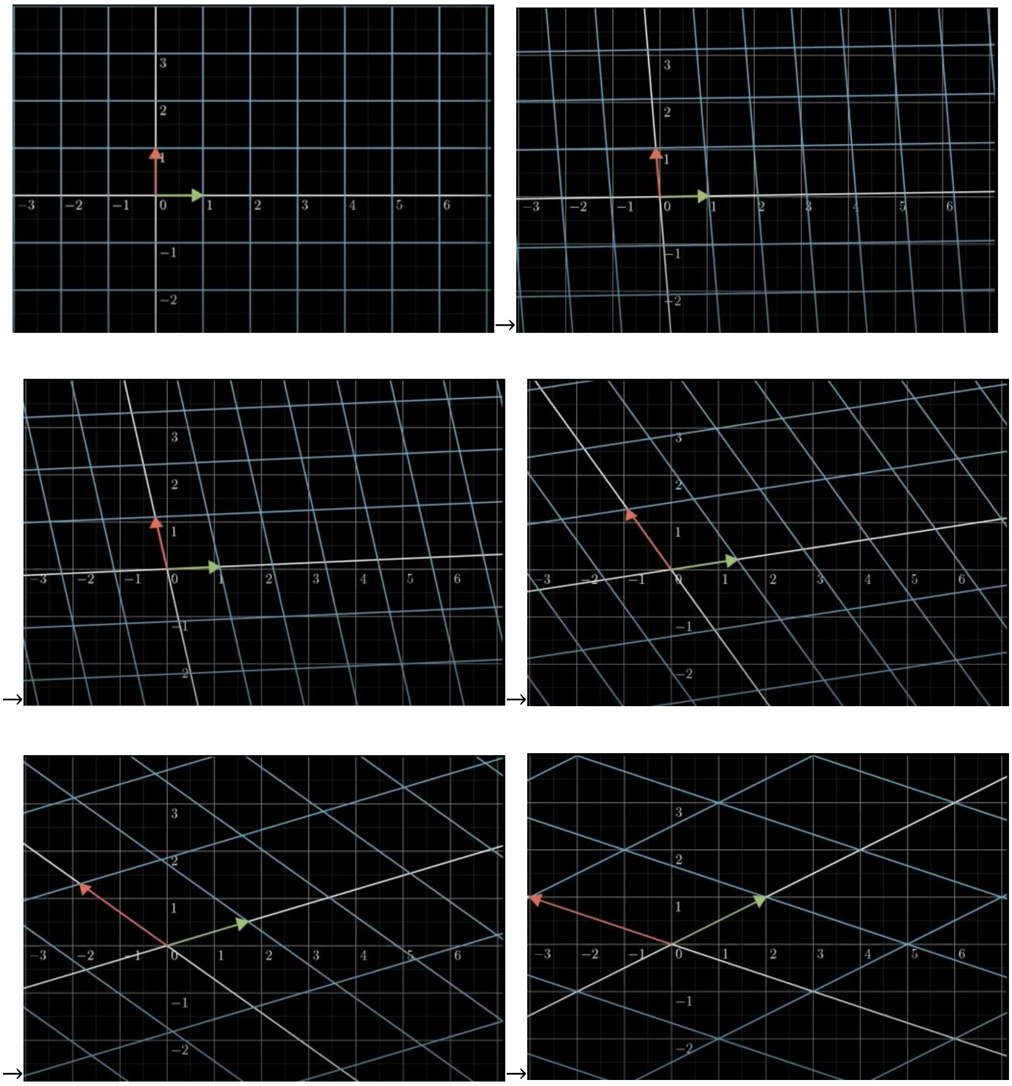
$$L(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = \alpha\, L(\mathbf{v}_1) + \beta\, L(\mathbf{v}_2)$$

for all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and for all scalars α and β.

For example,

$$\begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \to \begin{bmatrix} 2x + (-3)y \\ 1x + 1y \end{bmatrix}$$

So, we can show that from a vector space $V = \{(x,y)^T \,|\, x, y \in \mathrm{R}\}$ into a vector space $W = \{(u,v)^T | u = 2x + (-3)y, v = 1x + 1y,\ x, y \in \mathrm{R}\}$,



Note all of the grid lines remain parallel and evenly spaced.
From above figures, we can get

$$\text{Green vector} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{Red vector} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

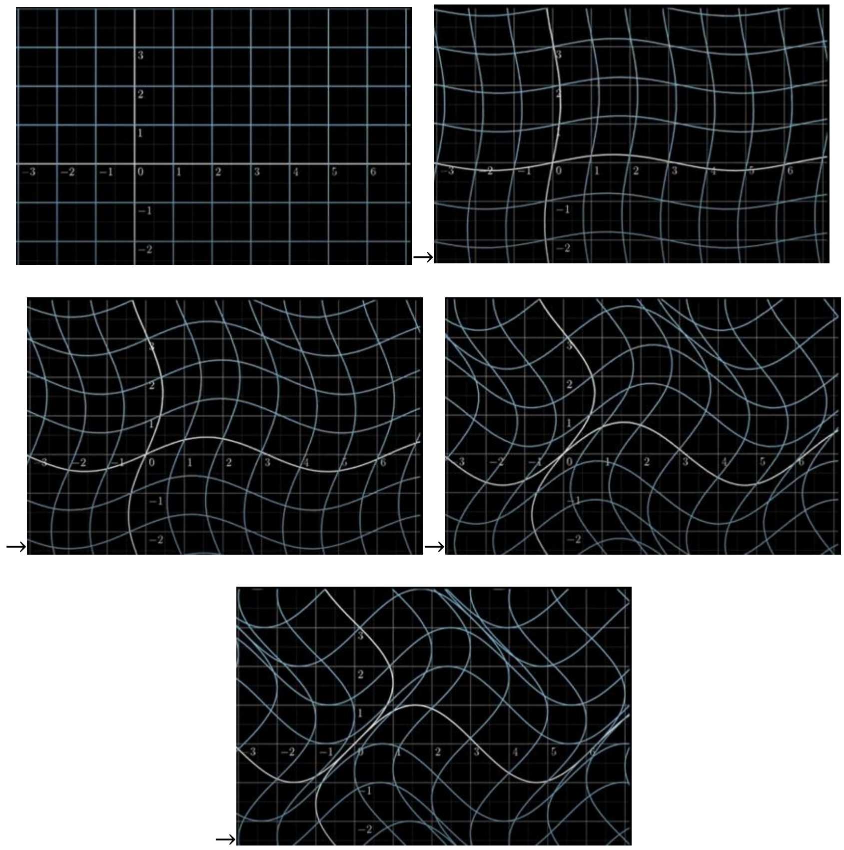
$$\begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

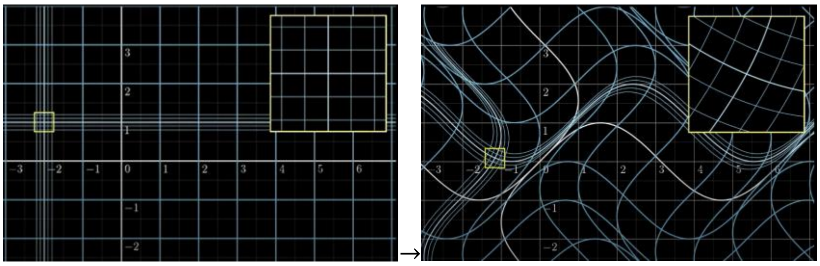
$$L\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = L\left(x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$$

Now, from linear algebra to multivariable function, for example, a nonlinear function,

$$f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + \sin(y) \\ y + \sin(x) \end{bmatrix}$$

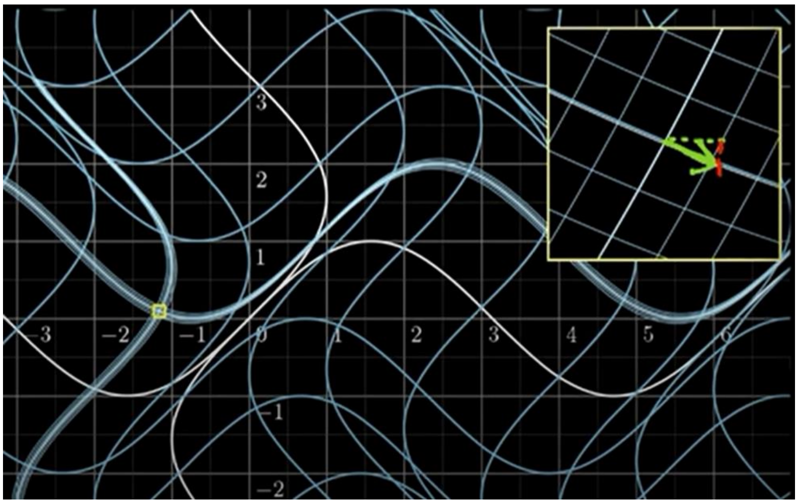


Here we will show the Local linearity for a multivariable function, by looking at the really small neighborhood of the point $\begin{bmatrix} x \\ y \end{bmatrix} = (-2,1)^T$,



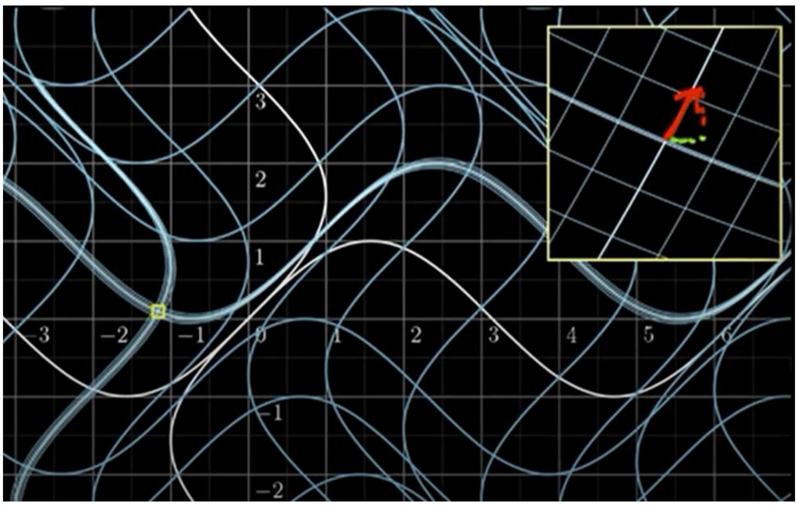
The neighborhood above looks precisely like a certain linear function. So, it reminds us to find the matrix to represent the linear function that this looks like.
For the tiny step at x-axis,

$$f\left(\begin{bmatrix} dx \\ 0 \end{bmatrix}\right) = \begin{bmatrix} dx + \sin(0) \\ 0 + \sin(dx) \end{bmatrix} = \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix} |_{(dx,0)^T} = \begin{bmatrix} f_1(dx,0) \\ f_2(dx,0) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \frac{\partial f_2}{\partial x} \end{bmatrix}$$



For the tiny step at y-axis,

$$f\left(\begin{bmatrix} 0 \\ dy \end{bmatrix}\right) = \begin{bmatrix} 0 + \sin(dy) \\ dy + \sin(0) \end{bmatrix} = \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix} |_{(0,dy)^T} = \begin{bmatrix} f_1(0,dy) \\ f_2(0,dy) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial y} \end{bmatrix}$$



The matrix,

$$\begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix}$$

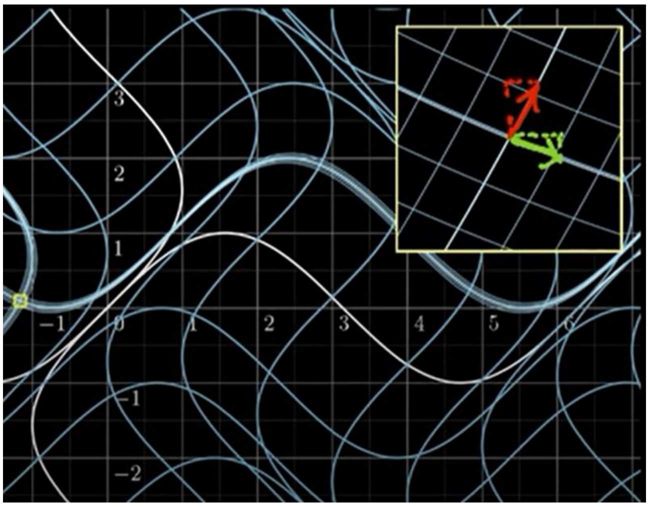
is called the **Jacobian matrix**.

For the point $(-2,1)$ at the nonlinear transformation $f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + \sin(y) \\ y + \sin(x) \end{bmatrix}$, the Jacobian matrix is

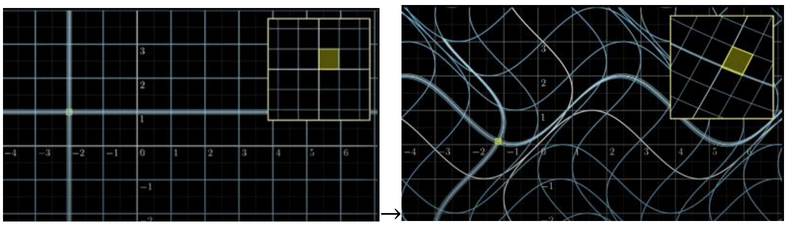
$$\begin{bmatrix} 1 & \cos(y) \\ \cos(x) & 1 \end{bmatrix}$$

At point $(-2,1)$, the Jacobian matrix is

$$\begin{bmatrix} 1 & \cos(1) \\ \cos(-2) & 1 \end{bmatrix} \approx \begin{bmatrix} 1 & 0.59 \\ -0.92 & 1 \end{bmatrix}$$



For the Jacobian Determinant, we illustrate the tiny area as below using same nonlinear transformation,



Assuming the left area of yellow square is 1, the determinant of the Jacobian matrix represents the factor by which it gets stretched to the right figure.

Now that we have the Jacobian out of the way, we can give the formula for change of variables for a double integral.

Change of Variables for a Double Integral

Suppose that we want to integrate $f(x,y)$ over the region R. Under the transformation $x = g(u,v)$, $y = h(u,v)$, the region becomes S and the integral becomes,

$$\iint\limits_R f(x,y)\,dA = \iint\limits_S f\big(g(u,v),h(u,v)\big)\left|\frac{\partial(x,y)}{\partial(u,v)}\right| d\bar{A}$$

Note that we use $d\bar{A}$ in the u/v integral above to denote that it will be in terms of du and dv, once we convert to two single integrals rather than the dx and dy we are used for dA. This is notational only and we generally use dA for both and make sure to remember that the "new" dA is in terms of du and dv.

Also note that we are taking the absolute value of the Jacobian.

If we look just at the differentials in the above formula, we can also say that

$$dA = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| d\bar{A}$$

Example 3: Show that when changing to polar coordinates we have $dA = r dr d\theta$,

The transformation here is the standard conversion formulas,

$$x = r\cos\theta \quad y = r\sin\theta$$

The Jacobian for this transformation is,

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix}$$

$$= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}$$

$$= r$$

We then get,

$$dA = \left|\frac{\partial(x,y)}{\partial(r,\theta)}\right| dr d\theta = |r| dr d\theta = r dr d\theta$$

Let's now briefly look at triple integrals. In this case we will again start with a region R and use the transformation $x = g(u,v,w)$, $y = h(u,v,w)$, and $z = k(u,v,w)$ to transform the region into the new region S. Here is the definition of the Jacobian for this kind of transformation.

$$\frac{\partial(x,y,z)}{\partial(u,v,w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix} = \det\begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{bmatrix}$$

In this case the Jacobian is defined in terms of the determinant of a 3×3 matrix.

The integral under this transformation is,

$$\iiint\limits_R f(x,y,z)\,dV = \iiint\limits_S f\big(g(u,v,w),h(u,v,w),k(u,v,w)\big)\left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| d\bar{V}$$

As with double integrals we used $d\bar{V}$, in the $u/v/w$ integral above we will need to use du, dv and dw when converting to single integrals.
We can look at the differentials and note that we must have

$$dV = \left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| d\bar{V}$$

Example 6: Verify that $dV = \rho^2 \sin\varphi \, d\rho d\theta d\varphi$ when using spherical coordinates.
Here the transformation is just the standard conversion formulas.

$$x = \rho\sin\varphi\cos\theta \quad y = \rho\sin\varphi\sin\theta \quad z = \rho\cos\varphi$$

The Jacobian is,

$$\frac{\partial(x,y,z)}{\partial(\rho,\theta,\varphi)} = \begin{vmatrix} \sin\varphi\cos\theta & -\rho\sin\varphi\sin\theta & \rho\cos\varphi\cos\theta \\ \sin\varphi\sin\theta & \rho\sin\varphi\cos\theta & \rho\cos\varphi\sin\theta \\ \cos\varphi & 0 & -\rho\sin\varphi \end{vmatrix}$$

$$= -\rho^2\sin\varphi$$

Finally, dV becomes,

$$dV = |-\rho^2\sin\varphi| d\rho d\theta d\varphi = \rho^2\sin\varphi \, d\rho d\theta d\varphi$$

Recall that we restricted φ to the range $0 \le \varphi \le \pi$ for spherical coordinates and so we know that $\sin\varphi \ge 0$, and so we don't need the absolute value bars on the sine.

ADDITION CHAPTER 4 Line Integrals

Vector Fields

Definition

A vector field on two (or three) dimensional space is a function $\vec{F}$ that assigns to each point (x, y) (or (x, y, z)) a two (or three dimensional) vector given by $\vec{F}(x, y)$ (or $\vec{F}(x, y, z)$).

That may not make a lot of sense, but most people do know what a vector field is, or at least they've seen a sketch of a vector field. If you've seen a current sketch, giving the direction and magnitude of a flow of a fluid or the direction and magnitude of the winds, then you've seen a sketch of a vector field.

The standard notation for the function $\vec{F}$ is,

$$\vec{F}(x, y) = P(x, y)\vec{\imath} + Q(x, y)\vec{\jmath}$$
$$\vec{F}(x, y, z) = P(x, y, z)\vec{\imath} + Q(x, y, z)\vec{\jmath} + R(x, y, z)\vec{k}$$

depending on whether or not we're in two or three dimensions. The function P, Q, R (if it is present) are sometimes called **scalar functions**.

Example 1: Sketch each of the following vector fields,

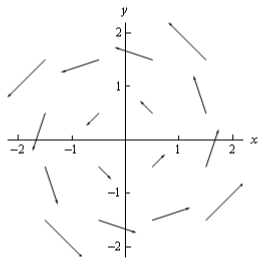
(a) $\vec{F}(x, y) = -y\vec{\imath} + x\vec{\jmath}$

Okay, to graph the vector field we need to get some "values" of the function. This means plugging in some points into the function. Here are a couple of evaluations.

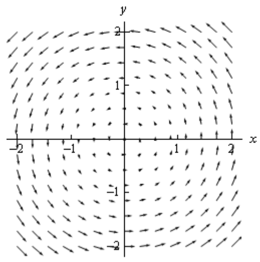
$$\vec{F}\left(\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2}\vec{\imath} + \frac{1}{2}\vec{\jmath}$$
$$\vec{F}\left(\frac{1}{2}, -\frac{1}{2}\right) = -\left(-\frac{1}{2}\right)\vec{\imath} + \frac{1}{2}\vec{\jmath} = \frac{1}{2}\vec{\imath} + \frac{1}{2}\vec{\jmath}$$
$$\vec{F}\left(\frac{3}{2}, \frac{1}{4}\right) = -\frac{1}{4}\vec{\imath} + \frac{3}{2}\vec{\jmath}$$

So, just what do these evaluations tell us? Well the first one tells us that at the point $\left(\frac{1}{2}, \frac{1}{2}\right)$ we will plot the vector $-\frac{1}{2}\vec{\imath} + \frac{1}{2}\vec{\jmath}$. Likewise, the third evaluation tells us that at the point $\left(\frac{3}{2}, \frac{1}{4}\right)$ we will plot the vector $-\frac{1}{4}\vec{\imath} + \frac{3}{2}\vec{\jmath}$.

We can continue in this fashion plotting vectors for several points, and we'll get the following sketch of the vector field.



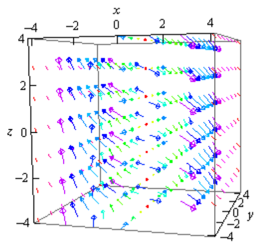
If we want significantly more points plotted, then it is usually best to use a computer aided graphing system such as Maple or Mathematica. Here is a sketch with many more vectors included that was generated with Mathematica.

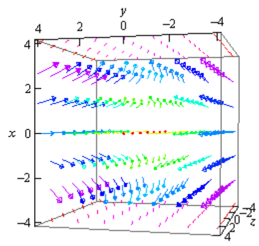


(b) $\vec{F}(x, y, z) = 2x\vec{\imath} - 2y\vec{\jmath} - 2x\vec{k}$

Notice that z only affects the placement of the vector in this case and does not affect the direction or the magnitude of the vector. Sometimes this will happen so don't get excited about it when it does.

Here is a couple of sketches generated by Mathematica. The sketch on the left is from the "front" and the sketch on the right is from "above".





Now that we've seen a couple of vector fields, let's notice that we've already seen a vector field function. In the second chapter we looked at the **gradient vector**. Recall that given a function $f(x, y, z)$, the gradient vector is defined by,

$$\nabla f = \langle f_x, f_y, f_z \rangle$$

This is a vector field and is often called a **gradient vector field**.

Let's do another example that will illustrate the relationship between the gradient vector field of a function and its contours.

Example 3: Sketch the gradient vector field for $f(x, y) = x^2 + y^2$ as well as several contours for this function:

Recall that the contours for a function are nothing more than curves defined by,

$$f(x, y) = k$$

for various values of k. So, for our function the contours are defined by the equation,

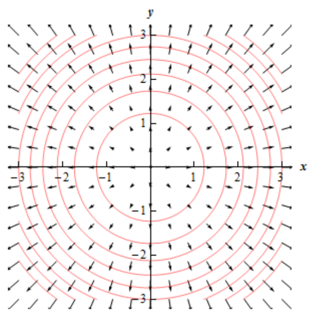
$$x^2 + y^2 = k$$

and so they are circles centered at the origin with radius $\sqrt{k}$.

Here is the gradient vector field for this function.

$$\nabla f\,(x, y) = 2x\vec{\imath} + 2y\vec{\jmath}$$

Here is a sketch of several of the contours as well as the gradient vector field.



Notice that the vectors of the vector field are all orthogonal (or perpendicular) to the contours. This will always be the case when we are dealing with the contours of a function as well as its gradient vector field.

The k's we used for the graph above were $1.5, 3, 4.5, 6, 7.5, 9, 10.5, 12$, and 13.5. Now notice that as we increased k by 1.5, the contour curves get closer together and as the contour curves get closer together the larger the vectors become. In other words, the closer the contour curves are (as k is increased by a fixed amount) the faster the function is changing at that point. Also recall that the direction of fastest change for a function is given by the gradient vector at that point. Therefore, it should make sense that the two ideas should match up as they do here.

The final topic of this section is the conservative vector fields. A vector field $\vec{F}$ is called a **conservative vector field** if there exists a function f such that $\vec{F} = \nabla f$. If $\vec{F}$ is a conservative vector field then the function, f, is called a **potential function** for $\vec{F}$.

All this definition is saying that a vector field is conservative if it is also a gradient vector field for some function.

Line Integrals - Part I

Here are some of the more basic curves that we'll need to know how to do, as well as limits on the parameter if they are required.

Curve	Parametric Equations	
$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ (Ellipse)	Counter-Clockwise $x = a\cos(t)$ $y = b\sin(t)$ $0 \le t \le 2\pi$	Clockwise $x = a\cos(t)$ $y = -b\sin(t)$ $0 \le t \le 2\pi$
$x^2 + y^2 = r^2$ (Circle)	Counter-Clockwise $x = r\cos(t)$ $y = r\sin(t)$ $0 \le t \le 2\pi$	Clockwise $x = r\cos(t)$ $y = -r\sin(t)$ $0 \le t \le 2\pi$
$y = f(x)$	$x = t$ $y = f(t)$	
$x = g(y)$	$x = g(t)$ $y = t$	
Line Segment From (x_0, y_0, z_0) to (x_1, y_1, z_1)	$\vec{r}(t) = (1-t)\langle x_0, y_0, z_0\rangle + t\langle x_1, y_1, z_1\rangle, 0 \le t \le 1$ or $x = (1-t)x_0 + tx_1$ $y = (1-t)y_0 + ty_1, \quad 0 \le t \le 1$ $z = (1-t)z_0 + tz_1$	

With line integrals we will start with integrating the function $f(x, y)$, a function of two variables, and the values of x and y that we're going to use will be the points, (x, y), that lie on a curve C. Note that this is different from the double integrals that we were working with in the previous chapter, where the points came out of some two-dimensional region.

Let's start with the curve C that the points come from. We will assume that the curve is smooth (defined shortly) and is given by the parametric equations,

$$x = h(t), \quad y = g(t), \quad a \le t \le b$$

We will often want to write the parameterization of the curve as a vector function. In this case the curve is given by,

$$\vec{r}(t) = h(t)\vec{i} + g(t)\vec{j}, \quad a \le t \le b$$

The curve is called **smooth** if $\vec{r}'(t)$ is continuous and $\vec{r}'(t) \neq 0$ for all t.

The **line integral** of $f(x,y)$ along C is denoted by,

$$\int_C f(x,y)ds$$

We use a ds here to acknowledge the fact that we are moving along the curve, C, instead of the x-axis (denoted by dx) or the y-axis (denoted by dy). Because of the ds, this is sometimes called the **line integral of f with respect to arc length**.

The ds is the same for both the arc length integral and the notation for the line integral. Recalling the arc length of a curve given by parametric equations, we found it to be,

$$L = \int_a^b ds, \quad \text{where } ds = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$$

So, to compute a line integral we will convert everything over to the parametric equations. The line integral is then,

$$\int_C f(x,y)ds = \int_a^b f\big(h(t), g(t)\big)\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$$

Don't forget to plug the parametric equations into the function as well.

If we use the vector form of the parameterization, we can simplify the notation up somewhat by noticing that,

$$\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} = \|\vec{r}'(t)\|$$

where $\|\vec{r}'(t)\|$ is the magnitude or norm of $\vec{r}'(t)$. Using this notation, the line integral becomes,

$$\int_C f(x,y)ds = \int_a^b f\big(h(t), g(t)\big)\|\vec{r}'(t)\|dt$$

Note that as long as the parameterization of the curve C is traced out exactly once as t increases from a to b, the value of the line integral will be independent of the parameterization of the curve.

Example 1: Evaluate $\int_C xy^4 ds$ where C is the right half of the circle, $x^2 + y^2 = 16$ traced out in a counter clockwise direction:

We first need a parameterization of the circle. This is given by,

$$x = 4\cos t, \quad y = 4\sin t$$

We now need a range of t's that will give the right half of the circle. The following range of t's will do this,

$$-\frac{\pi}{2} \le t \le \frac{\pi}{2}$$

Now, we need the derivatives of the parametric equations and let's compute ds.

$$\frac{dx}{dt} = -4\sin t, \quad \frac{dy}{dt} = 4\cos t$$

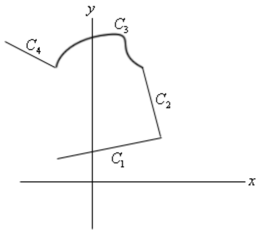
$$ds = \sqrt{16\sin^2 t + 16\cos^2 t}\, dt = 4dt$$

The line integral is then,

$$\int_C xy^4 ds = \int_{-\pi/2}^{\pi/2} 4\cos t\,(4\sin t)^4(4)dt$$

$$= \frac{8192}{5}$$

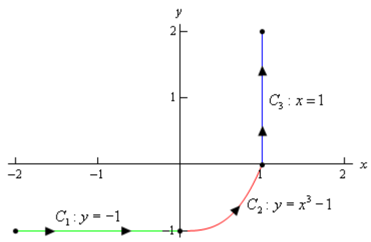
Next, we need to talk about line integrals over **piecewise smooth curves**. A piecewise smooth curve is any curve that can be written as the union of a finite number of smooth curves, C_1, $\cdots$, C_n where the end point of C_i is the starting point of C_{i+1}. Below is an illustration of a piecewise smooth curve.



Evaluation of line integrals over piecewise smooth curves is a relatively simple thing to do. All we do is evaluate the line integral over each of the pieces and then add them up. The line integral for some function over the above piecewise curve would be,

$$\int_C f(x,y)ds = \int_{C_1} f(x,y)ds + \int_{C_2} f(x,y)ds + \int_{C_3} f(x,y)ds + \int_{C_4} f(x,y)ds$$

Example 2: Evaluate $\int_C 4x^3 ds$ where C is the curve shown below:



So, first we need to parameterize each of the curves.

$$C_1: x = t, y = -1, \quad -2 \le t \le 0$$
$$C_2: x = t, y = t^3 - 1, \quad 0 \le t \le 1$$
$$C_3: x = 1, y = t, \quad 0 \le t \le 2$$

$$\int_C 4x^3 ds = \int_{C_1} 4x^3 ds + \int_{C_2} 4x^3 ds + \int_{C_3} 4x^3 ds$$

$$= -5.732.$$

Notice that we put direction arrows on the curve in the above example. The direction of motion along a curve may change the value of the line integral as we will see in the next section. Also note that the curve can be thought of a curve that takes us from $(-2,-1)$ to the point $(1,2)$.

Example 4: Evaluate $\int_C 4x^3 ds$ where C is the line segment from $(1,2)$ to $(-2,-1)$,

This one isn't much different, work wise, from the previous example. Here is the parameterization of the curve.

$$\begin{aligned}\vec{r}(t) &= (1-t)\langle 1,2\rangle + t\langle -2,-1\rangle \\ &= \langle 1-3t, 2-3t\rangle\end{aligned}$$

for $0 \le t \le 1$. Remember that we are switching the direction of the curve and this will also change the parameterization, so we can make sure that we start/end at the proper point.

$$\int_C 4x^3 ds = -15\sqrt{2}.$$

Let's suppose that the curve C has the parameterization $x = h(t)$, $y = g(t)$. Let's also suppose that the initial point on the curve is A, and the final point on the curve is B. The parameterization $x = h(t)$, $y = g(t)$ will then determine an **orientation** for the curve where the positive direction is the direction that is traced out as t increases. Finally, let $-C$ be the curve with the same points as C, however in this case the curve has B as the initial point and A as the final point, again t is increasing as we traverse this curve. In other words, given a curve C, the curve $-C$ is the same curve as C except the direction has been reversed.

Fact

$$\boxed{\int_C f(x,y)ds = \int_{-C} f(x,y)ds}$$

So, for a line integral with respect to arc length, we can change the direction of the curve and not change the value of the integral. This is a useful fact to remember as some line integrals will be easier in one direction than the other.

Let's suppose that the three-dimensional curve C is given by the parameterization,

$$x = x(t), \quad y = y(t), \quad z = z(t), \quad a \le t \le b$$

then the line integral is given by,

$$\boxed{\int_C f(x,y,z)ds = \int_a^b f\big(x(t),y(t),z(t)\big)\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt}$$

Note that often when dealing with three-dimensional space, the parameterization will be given as a vector function.

$$\vec{r}(t) = \langle x(t), y(t), z(t)\rangle$$

Also notice that, as with two-dimensional curves, we have,

$$\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + +\left(\frac{dz}{dt}\right)^2} = \|\vec{r}\,'(t)\|$$

and the line integral can again be written as,

$$\int_C f(x,y,z)ds = \int_a^b f\big(x(t),y(t),z(t)\big)\|\vec{r}\,'(t)\|dt$$

Line Integrals - Part II

In this section we want to look at line integrals with respect to x and/or y.

As with the last section, we will start with a two-dimensional curve C with parameterization,

$$x = x(t), \quad y = y(t), \quad a \le t \le b$$

The **line integral of f with respect to x** is,

$$\int_C f(x,y)dx = \int_a^b f\big(x(t),y(t)\big)x'(t)dt$$

The **line integral of f with respect to y** is,

$$\int_C f(x,y)dy = \int_a^b f\big(x(t),y(t)\big)y'(t)dt$$

Note that the only notational difference between these two and the line integral with respect to arc length (from the previous section) is the differential.

These two integrals often appear together and so we have the following shorthand notation for these cases.

$$\int_C Pdx + Qdy = \int_C P(x,y)dx + \int_C Q(x,y)dy$$

Example 1: Evaluate $\int_C \sin(\pi y)\,dy + yx^2dx$ where C is the line segment from $(0,2)$ to $(1,4)$:

Here is the parameterization of the curve,

$$\vec{r}(t) = (1-t)\langle 0,2\rangle + t\langle 1,4\rangle = \langle t, 2+2t\rangle, \quad 0 \le t \le 1$$

$$\int_C \sin(\pi y)\,dy + yx^2dx = \frac{7}{6}.$$

Example 2: Evaluate $\int_C \sin(\pi y)\,dy + yx^2dx$ where C is the line segment from $(1,4)$ to $(0,2)$:

Here is the new parameterization.

$$\vec{r}(t) = (1-t)\langle 1,4\rangle + t\langle 0,2\rangle = \langle 1-t, 4-2t\rangle, \quad 0 \le t \le 1$$

$$\int_C \sin(\pi y)\,dy + yx^2dx = -\frac{7}{6}$$

So, switching the direction of the curve got us a different value, or at least the opposite sign of the value from the first example. In fact, this will always happen with these kinds of line integrals.

Fact

If C is any curve then,

$$\int_{-C} f(x,y)dx = -\int_{C} f(x,y)dx \quad \text{and} \quad \int_{-C} f(x,y)dy = -\int_{C} f(x,y)dy$$

With the combined form of these two integrals we get,

$$\int_{-C} Pdx + Qdy = -\int_{C} Pdx + Qdy$$

We can also do these integrals over three-dimensional curves as well.

$$\int_{C} f(x,y,z)dx = \int_a^b f\big(x(t),y(t),z(t)\big)x'(t)dt$$

$$\int_{C} f(x,y,z)dy = \int_a^b f\big(x(t),y(t),z(t)\big)y'(t)dt$$

$$\int_{C} f(x,y,z)dz = \int_a^b f\big(x(t),y(t),z(t)\big)z'(t)dt$$

where the curve C is parameterized by

$$x = x(t), \quad y = y(t), \quad z = z(t), \quad a \le t \le b$$

These three will often occur together so the shorthand we'll be using here is,

$$\int_{C} Pdx + Qdy + Rdz = \int_{C} P(x,y,z)dx + \int_{C} Q(x,y,z)dy + \int_{C} R(x,y,z)dz$$

Line Integrals of Vector Fields

In this section we are going to evaluate line integrals of vector fields. We'll start with the vector field,

$$\vec{F}(x,y,z) = P(x,y,z)\vec{\imath} + Q(x,y,z)\vec{\jmath} + R(x,y,z)\vec{k}$$

and the three-dimensional, smooth curve given by

$$\vec{r}(t) = x(t)\vec{\imath} + y(t)\vec{\jmath} + z(t)\vec{k}, \quad a \le t \le b$$

The **line integral of $\vec{F}$ along C** is

$$\int_{C} \vec{F} \cdot d\vec{r} = \int_a^b \vec{F}\big(\vec{r}(t)\big) \cdot \vec{r}'(t)dt$$

Note the notation in the integral on the left side. That really is a **dot product** of the vector field and the differential really is a vector. Also, $\vec{F}\big(\vec{r}(t)\big)$ is a shorthand for,

$$\vec{F}\big(\vec{r}(t)\big) = \vec{F}\big(x(t),y(t),z(t)\big)$$

We can also write line integrals of vector fields as a line integral with respect to arc length as follows,

$$\int_C \vec{F} \cdot d\vec{r} = \int_C \vec{F} \cdot \vec{T} ds$$

where $\vec{T}(t)$ is the unit tangent vector and is given by,

$$\vec{T}(t) = \frac{\vec{r}'(t)}{\|\vec{r}'(t)\|}$$

If we use our knowledge on how to compute line integrals with respect to arc length, we can see that this second form is equivalent to the first form given above.

$$\int_C \vec{F} \cdot d\vec{r} = \int_C \vec{F} \cdot \vec{T} ds$$

$$= \int_a^b \vec{F}(\vec{r}(t)) \cdot \frac{\vec{r}'(t)}{\|\vec{r}'(t)\|} \|\vec{r}'(t)\| dt$$

$$= \int_a^b \vec{F}(\vec{r}(t)) \cdot \vec{r}'(t) dt$$

In general, we use the first form to compute these line integrals as it is usually much easier to use.

Example 1: Evaluate $\int_C \vec{F} \cdot d\vec{r}$ where $\vec{F}(x, y, z) = 8x^2yz\vec{i} + 5z\vec{j} - 4xy\vec{k}$ and C is the curve given by $\vec{r}(t) = t\vec{i} + t^2\vec{j} + t^3\vec{k}$, $0 \le t \le 1$,

$$\vec{F}(\vec{r}(t)) = 8t^2(t^2)(t^3)\vec{i} + 5t^3\vec{j} - 4t(t^2)\vec{k} = 8t^7\vec{i} + 5t^3\vec{j} - 4t^3\vec{k}$$

$$\vec{r}'(t) = \vec{i} + 2t\vec{j} + 3t^2\vec{k}$$

$$\vec{F}(\vec{r}(t)) \cdot \vec{r}'(t) = 8t^7 + 10t^4 - 12t^5$$

$$\int_C \vec{F} \cdot d\vec{r} = \int_0^1 8t^7 + 10t^4 - 12t^5 dt$$

$$= 1.$$

Given the vector field $\vec{F}(x, y, z) = P\vec{i} + Q\vec{j} + R\vec{k}$ and the curve C parameterized by $\vec{r}(t) = x(t)\vec{i} + y(t)\vec{j} + z(t)\vec{k}$, $a \le t \le b$ the line integral is,

$$\int_C \vec{F} \cdot d\vec{r} = \int_a^b (P\vec{i} + Q\vec{j} + R\vec{k}) \cdot (x'\vec{i} + y'\vec{j} + z'\vec{k}) dt$$

$$= \int_a^b (Px' + Qy' + Rz') dt = \int_a^b Px' dt + \int_a^b Qy' dt + \int_a^b Rz' dt = \int_C Pdx + \int_C Qdy + \int_C Rdz$$

$$= \int_C Pdx + Qdy + Rdz$$

$$\boxed{\int_C \vec{F} \cdot d\vec{r} = \int_C Pdx + Qdy + Rdz}$$

Note that this gives us another method for evaluating line integrals of vector fields.

This also allows us to say the following, about reversing the direction of the path with line integrals of vector fields.

Fact

$$\int_{-C} \vec{F} \cdot d\vec{r} = -\int_{C} \vec{F} \cdot d\vec{r}$$

Fundamental Theorem for Line Integrals

Theorem

Suppose that C is a smooth curve given by $\vec{r}(t)$, $a \le t \le b$. Also suppose that f is a function whose gradient vector, ∇f, is continuous on C. Then,

$$\int_{C} \nabla f \cdot d\vec{r} = f\big(\vec{r}(b)\big) - f\big(\vec{r}(a)\big)$$

Note that $\vec{r}(a)$ represents the initial point on C while $\vec{r}(b)$ represents the final point on C. The theorem will hold regardless of the number of variables in the function.

Proof

For the purposes of the proof, we'll assume that we're working in three-dimensions, but it can be done in any dimension.

$$\int_{C} \nabla f \cdot d\vec{r} = \int_a^b \nabla f\left(\vec{r}(t)\right) \cdot \vec{r}\,'(t)dt$$

$$= \int_a^b \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}\right) dt$$

Now, at this point we can use the Chain Rule to simplify the integrand as follows,

$$\int_a^b \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}\right) dt = \int_a^b \frac{d}{dt}\left[f\big(\vec{r}(t)\big)\right]dt$$

To finish this off we just need to use the Fundamental Theorem of Calculus for single integrals.

$$\int_{C} \nabla f \cdot d\vec{r} = f\big(\vec{r}(b)\big) - f\big(\vec{r}(a)\big).$$

Example 1: Evaluate $\int_C \nabla f \cdot d\vec{r}$ where $f(x,y,z) = \cos(\pi x) + \sin(\pi y) - xyz$ and C is *any* path that starts at $\left(1,\frac{1}{2},2\right)$ and ends at $(2,1,-1)$,

The theorem above tells us that all we need are the initial and final points on the curve, in order to evaluate this kind of line integral. So, let $\vec{r}(t)$, $a \le t \le b$ be any path that starts at $\left(1,\frac{1}{2},2\right)$ and ends at $(2,1,-1)$. Then,

$$\vec{r}(a) = \langle 1,\frac{1}{2},2\rangle, \quad \vec{r}(b) = \langle 2,1,-1\rangle$$

The integral is then,

$$\int_{C} \nabla f \cdot d\vec{r} = f(2,1,-1) - f\left(1,\frac{1}{2},2\right)$$

$$= 4.$$

The important idea from this example (and hence about the Fundamental Theorem of Calculus) is that, for these kinds of line integrals, we didn't really need to know the path to get the answer. In other words, we could use any path we want and we'll always get the same results.

We now have a type of line integral for which we know that changing the path will NOT change the value of the line integral.

Here are some definitions.

Definitions

First suppose that $\vec{F}$ is a continuous vector field in some domain D.

1. $\vec{F}$ is a **conservative** vector field if there is a function f such that $\vec{F} = \nabla f$. The function f is called a **potential function** for the vector field.

2. $\int_C \vec{F} \cdot d\vec{r}$ is **independent of path** if $\int_{C_1} \vec{F} \cdot d\vec{r} = \int_{C_2} \vec{F} \cdot d\vec{r}$ for any two paths C_1 and C_2 in D with the same initial and final points.

3. A path C is called **closed** if its initial and final points are the same point. For example, a circle is a closed path.

4. A path C is **simple** if it doesn't cross itself. A circle is a simple curve while a figure 8 type curve is not simple.

5. A region D is **open** if it doesn't contain any of its boundary points.

6. A region D is **connected** if we can connect any two points in the region with a path that lies completely in D

7. A region D is **simply-connected** if it is connected and it contains no holes.

With these definitions we can now give some nice facts.

Facts

1. $\int_C \nabla f \cdot d\vec{r}$ is independent of path.

2. If $\vec{F}$ is a conservative vector field then $\int_C \vec{F} \cdot d\vec{r}$ is independent of path. $(\int_C \vec{F} \cdot d\vec{r} = \int_C \nabla f \cdot d\vec{r})$

3. If $\vec{F}$ is a continuous vector field on an open connected region D and if $\int_C \vec{F} \cdot d\vec{r}$ is independent of path (for any path in D) then $\vec{F}$ is a conservative vector field on D.

4. If $\int_C \vec{F} \cdot d\vec{r}$ is independent of path then $\int_C \vec{F} \cdot d\vec{r} = 0$ for every closed path C.

5. If $\int_C \vec{F} \cdot d\vec{r} = 0$ for every closed path C then $\int_C \vec{F} \cdot d\vec{r}$ is independent of path.

Conservative Vector Fields

In this section we want to look at two questions. First, given a vector field $\vec{F}$, is there any way of determining if it is a conservative vector field? Secondly, if we know that $\vec{F}$ is a conservative vector field, how do we go about finding a potential function for the vector field?

The first question is easy to answer at this point if we have a two-dimensional vector field. For higher dimensional vector fields, we'll need to wait until the final section in this chapter to answer this question.

Theorem

Let $\vec{F} = P\vec{\imath} + Q\vec{\jmath}$ be a vector field on an **open** and **simply-connected** region D. Then if P and Q have continuous first order partial derivatives in D and

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$$

the vector field $\vec{F}$ is conservative.

Now that we know how to identify if a two-dimensional vector field is conservative, we need to address how to find a potential function for the vector field. This is actually a fairly simple process.

Since,

$$\nabla f = \frac{\partial f}{\partial x}\vec{\imath} + \frac{\partial f}{\partial y}\vec{\jmath} = P\vec{\imath} + Q\vec{\jmath} = \vec{F}$$

by setting components equal we have,

$$\frac{\partial f}{\partial x} = P \quad \text{and} \quad \frac{\partial f}{\partial y} = Q$$

By integrating each of these with respect to the appropriate variable, we can arrive at the following two equations.

$$f(x,y) = \int P(x,y)dx \quad \text{or} \quad f(x,y) = \int Q(x,y)dy$$

Example 2: Determine if the following vector fields are conservative and find a potential function for the vector field if it is conservative,

(a) $\vec{F} = (2x^3y^4 + x)\vec{\imath} + (2x^4y^3 + y)\vec{\jmath}$

Let's first identify P and Q and then check that the vector field is conservative.

$$P = 2x^3y^4 + x, \quad \frac{\partial P}{\partial y} = 8x^3y^3$$

$$Q = 2x^4y^3 + y, \quad \frac{\partial Q}{\partial x} = 8x^3y^3$$

Now let's find the potential function.

$$\frac{\partial f}{\partial x} = 2x^3y^4 + x \quad \text{and} \quad \frac{\partial f}{\partial y} = 2x^4y^3 + y$$

From these we can see that

$$f(x,y) = \int 2x^3y^4 + xdx \quad \text{or} \quad f(x,y) = \int 2x^4y^3 + ydy$$

Here is the first integral.

$$f(x,y) = \int 2x^3y^4 + xdx$$

$$= \frac{1}{2}x^4y^4 + \frac{1}{2}x^2 + h(y)$$

where $h(y)$ is the "constant of integration".

We now need to determine $h(y)$.

$$\frac{\partial f}{\partial y} = \frac{\partial\left(\frac{1}{2}x^4y^4 + \frac{1}{2}x^2 + h(y)\right)}{\partial y} = 2x^4y^3 + h'(y) = 2x^4y^3 + y = Q$$

From this we can see that,

$$h'(y) = y$$

Notice that since $h'(y)$ is a function only of y, so if there are any x's in the equation at this point we will know that we've made a mistake.

$$h(y) = \int h'(y)dy = \frac{1}{2}y^2 + c$$

So, putting this all together we can see that a potential function for the vector field is,

$$f(x, y) = \frac{1}{2}x^4y^4 + \frac{1}{2}x^2 + \frac{1}{2}y^2 + c.$$

Now, as noted above, we don't have a way (yet) of determining if a three-dimensional vector field is conservative or not. However, if we are given that a three-dimensional vector field is conservative, finding a potential function is similar to the above process, although the work will be a little more involved.

In this case we will use the fact that,

$$\nabla f = \frac{\partial f}{\partial x}\vec{\imath} + \frac{\partial f}{\partial y}\vec{\jmath} + \frac{\partial f}{\partial z}\vec{k} = P\vec{\imath} + Q\vec{\jmath} + R\vec{k} = \vec{F}$$

Example 5: Evaluate $\int_C \vec{F} \cdot d\vec{r}$ where $\vec{F} = (2x^3y^4 + x)\vec{\imath} + (2x^4y^3 + y)\vec{\jmath}$ and C is given by

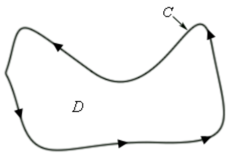
$\vec{r}(t) = (t\cos(\pi t) - 1)\vec{\imath} + \sin\left(\frac{\pi t}{2}\right)\vec{\jmath},\ 0 \le t \le 1,$

We could use the techniques we discussed when we first looked at line integrals of vector fields, or, take advantage of the fact that we know from above this vector field is conservative.

Green's Theorem

In this section we are going to investigate the relationship between certain kinds of line integrals (on closed paths) and double integrals.

Let's start off with a **simple** (recall that this means that it doesn't cross itself) **closed** curve C and let D be the region enclosed by the curve.



First, notice that because the curve is simple and closed, there are no holes in the region D. Also notice that a direction has been put on the curve. We will use the convention here that the curve C has a **positive orientation** if it is traced out in a counter-clockwise direction. Another way to think of a positive orientation (that will cover much more general curves as well see later) is that as we traverse the path following the positive orientation, the region D must always be on the **left**.

Given curves/regions such as this we have the following theorem.

Green's Theorem

Let C be a positively oriented, piecewise smooth, simple, closed curve and let D be the region enclosed by the curve. If P and Q have continuous first order partial derivatives on D then,

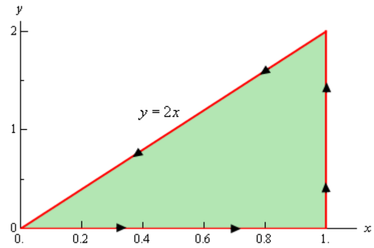
$$\int_C Pdx + Qdy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA$$

When working with a line integral in which the path satisfies the condition of Green's Theorem, we will often denote the line integral as,

$$\oint_C Pdx + Qdy \quad \text{or} \quad \oint_C Pdx + Qdy$$

Example 1: Use Green's Theorem to evaluate $\oint_C xydx + x^2y^3dy$ where C is the triangle with vertices $(0,0)$, $(1,0)$, $(1,2)$ with positive orientation,

Let's first sketch C and D for this case to make sure that the conditions of Green's Theorem are met for C, and will need the sketch of D to evaluate the double integral.



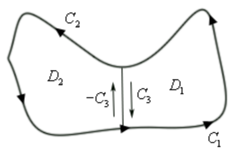
We can see that the following inequalities will define the region enclosed.

$$0 \le x \le 1, \quad 0 \le y \le 2x$$
$$P = xy, \quad Q = x^2y^3$$

$$\oint_C xydx + x^2y^3dy = \iint_D 2xy^3 - xdA$$

$$= \frac{2}{3}.$$

Let's start with the following region.



The region D will be $D_1 \cup D_2$. The boundary of D_1 is $C_1 \cup C_3$ while the boundary of D_2 is $C_2 \cup (-C_3)$. We can think of the whole boundary, C, as,

$$C = (C_1 \cup C_3) \cup \big(C_2 \cup (-C_3)\big) = C_1 \cup C_2$$

$$\iint_D (Q_x - P_y)dA = \iint_{D_1 \cup D_2} (Q_x - P_y)dA = \iint_{D_1} (Q_x - P_y)dA + \iint_{D_2} (Q_x - P_y)dA$$

Using Green's theorem,

$$\iint_{D_1} (Q_x - P_y) dA + \iint_{D_2} (Q_x - P_y) dA = \oint_{C_1 \cup C_3} Pdx + Qdy + \oint_{C_2 \cup (-C_3)} Pdx + Qdy$$

$$= \oint_{C_1} Pdx + Qdy + \oint_{C_3} Pdx + Qdy + \oint_{C_2} Pdx + Qdy + \oint_{-C_3} Pdx + Qdy$$

$$= \oint_{C_1} Pdx + Qdy + \oint_{C_2} Pdx + Qdy$$

Finally,

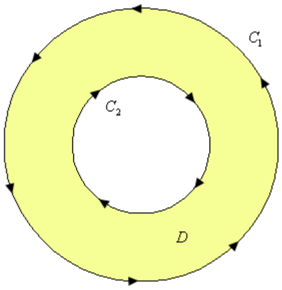
$$\iint_D (Q_x - P_y) dA = \oint_{C_1} Pdx + Qdy + \oint_{C_2} Pdx + Qdy$$

$$= \oint_{C_1 \cup C_2} Pdx + Qdy$$

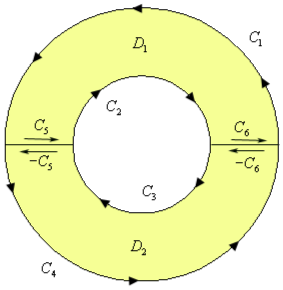
$$= \oint_C Pdx + Qdy$$

So, if we break a region up as we did above, then the portion of the line integral on the pieces of the curve that are in the middle of the region (each of which are in the opposite direction) will cancel out. This idea will help us in dealing with regions that have holes in them.

To see this let's look at a ring.



Now, since this region has a hole in it, we will apparently not be able to use Green's Theorem on any line integral with the curve $C = C_1 \cup C_2$. However, if we cut the disk in half and rename all the various portions of the curves, we get the following sketch.



We can do the following,

$$\iint_D (Q_x - P_y) dA = \iint_{D_1} (Q_x - P_y) dA + \iint_{D_2} (Q_x - P_y) dA$$

$$= \oint_{C_1 \cup C_2 \cup C_5 \cup C_6} Pdx + Qdy + \oint_{C_3 \cup C_4 \cup (-C_5) \cup (-C_6)} Pdx + Qdy$$

$$= \oint_{C_1} Pdx + Qdy + \oint_{C_2} Pdx + Qdy + \oint_{C_3} Pdx + Qdy + \oint_{C_4} Pdx + Qdy$$

$$= \oint_{C_1 \cup C_2 \cup C_3 \cup C_4} Pdx + Qdy$$

$$= \oint_{C} Pdx + Qdy$$

The end result of all of this is that we could have just used Green's Theorem on the disk from the start, even though there is a hole in it. This will be true in general for regions that have holes in them.