Math 132 Precalculus Review

Algebra

1. Whole numbers. Division with remainder: $a \div b = q$ rem r means $\frac{a}{b} = q + \frac{r}{b}$

2. Signs: $-(-a) = a,\ (-a)b = -ab,\ (-a)(-b) = ab,\ a - b = a + (-b)$

3. Fractions:
$$\frac{a}{b} \pm \frac{c}{d} = \frac{ad}{bd} \pm \frac{bc}{bd} = \frac{ad \pm bc}{bd}$$
$$\frac{a}{b}\cdot\frac{c}{d} = \frac{ac}{bd} \qquad \frac{1}{c/d} = \frac{d}{c} \qquad \frac{a/b}{c/d} = \frac{a}{b}\cdot\frac{d}{c} = \frac{ad}{bc}$$

4. Distributive law: $a(b+c) = ab + ac$
Hence: $\frac{b+c}{a} = \frac{1}{a}(b+c) = \frac{b}{a} + \frac{c}{a},\quad -(b+c) = (-1)(b+c) = -b-c$
Mistakes: $\sqrt{x^2+9} \neq x+3,\quad (\sqrt{x}+3)^2 \neq x+9,\quad \frac{1}{x+3} \neq \frac{1}{x} + \frac{1}{3}$

5. Powers: $a^n = a\cdots a$ (n factors), $a^n a^m = a^{n+m},\ (ab)^n = a^n b^n,\ (a^n)^m = a^{nm}$
$a^{-n} = \frac{1}{a^n},\quad a^{1/n} = \sqrt[n]{a},\quad a^{m/n} = \sqrt[n]{a^m},\quad a^{-m/n} = \frac{1}{\sqrt[n]{a^m}}$

6. Difference of powers: $a^2 - b^2 = (a-b)(a+b),\quad a^3 - b^3 = (a-b)(a^2+ab+b^2)$
$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \cdots + b^{n-1})$$
Sum of geometric series: $1 + r + r^2 + \cdots + r^{n-1} = \frac{1-r^n}{1-r}$ (take $a=1$, $b=r$)

7. Binomial powers: $(a \pm b)^2 = a^2 \pm 2ab + b^2,\quad (a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$
Binomial Theorem:
$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n-1}ab^{n-1} + b^n,$$
where $\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \ldots, \binom{n}{n-1}, \binom{n}{n}$ is the row $1, n, \ldots, n, 1$ in Pascal's Triangle:
$$\begin{array}{ccccccccccccc} &&&&&&1&&&&&& \\ &&&&&1&&1&&&&& \\ &&&&1&&2&&1&&&& \\ &&&1&&3&&3&&1&&& \\ &&1&&4&&6&&4&&1&& \\ &1&&5&&10&&10&&5&&1& \\ 1&&6&&15&&20&&15&&6&&1 \end{array}$$

8. Rationalize the denominator: $\frac{1}{\sqrt{a}} = \frac{1}{\sqrt{a}}\cdot\frac{\sqrt{a}}{\sqrt{a}} = \frac{\sqrt{a}}{a}$
$$\frac{1}{\sqrt{a}+\sqrt{b}} = \frac{1}{\sqrt{a}+\sqrt{b}}\cdot\frac{\sqrt{a}-\sqrt{b}}{\sqrt{a}-\sqrt{b}} = \frac{\sqrt{a}-\sqrt{b}}{\sqrt{a}^2-\sqrt{b}^2} = \frac{\sqrt{a}-\sqrt{b}}{a-b}$$

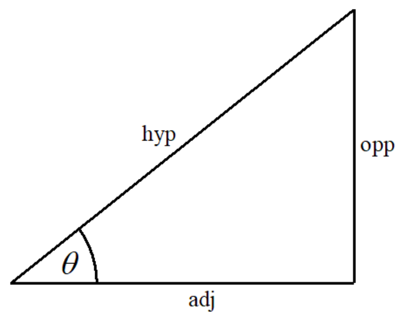
9. Quadratic formula: Gives the roots of a quadratic equation.
$$ax^2 + bx + c = 0 \qquad \Longleftrightarrow \qquad x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}.$$
This is not a real number when $b^2 - 4ac < 0$: there is no real solution $x \in \mathbb{R}$.

Notes by Peter Magyar `magyar@math.msu.edu`

Trigonometry

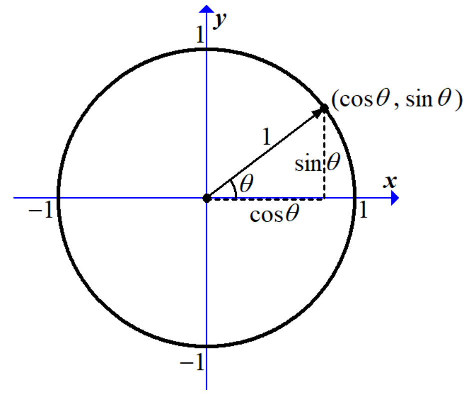
1. Right triangle definitions

$$\sin\theta = \frac{\text{opp}}{\text{hyp}} \qquad \cos\theta = \frac{\text{adj}}{\text{hyp}} \qquad \tan\theta = \frac{\text{opp}}{\text{adj}}$$



2. Circular coordinates

$$\cos^2\theta + \sin^2\theta = 1$$



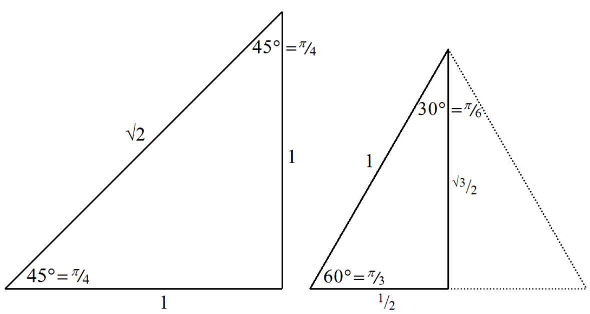
3. Quotient and co-functions

$$\tan\theta = \frac{\sin\theta}{\cos\theta} \qquad \cot\theta = \frac{\cos\theta}{\sin\theta} \qquad \sec\theta = \frac{1}{\cos\theta} \qquad \csc\theta = \frac{1}{\sin\theta}$$

$$\cos\theta = \sin(90°-\theta) \qquad \cot\theta = \tan(90°-\theta) \qquad \csc\theta = \sec(90°-\theta)$$

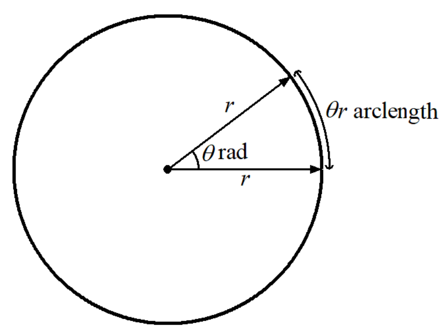
4. Special Triangles

isosceles right half equilateral



5. Radian angle measure

$360° = 2\pi$ rad $1° = \frac{2\pi}{360}$ rad



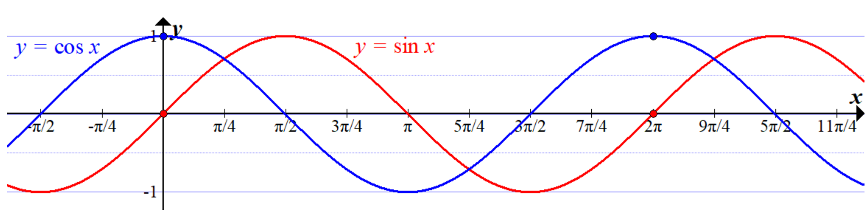
6. Angle addition formulas

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta \qquad \cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$

$$\sin(2\alpha) = 2\sin\alpha\cos\alpha \qquad \cos(2\alpha) = \cos^2\alpha - \sin^2\alpha = 1 - 2\sin^2\alpha = 2\cos^2\alpha - 1$$

7. Half angle formulas $\sin\left(\frac{\theta}{2}\right) = \sqrt{\frac{1-\cos\theta}{2}}$ $\cos\left(\frac{\theta}{2}\right) = \sqrt{\frac{1+\cos\theta}{2}}$

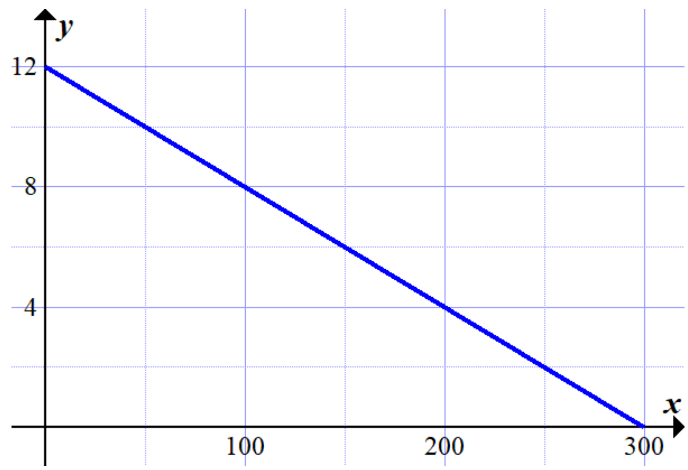
8. Graphs



Linear function example: gallons per mile

1. Physical: A gas tank has y gallons after x miles driven. Start at 12 gal, with constant gas usage rate 25 mi/gal, or $\frac{1}{25}$ gal/mi.
   Question: How far can the car travel until it runs out of gas?

2. Graphical: Linear function $y = f(x)$, y-intercept $(x, y) = (0, 12)$, slope $-\frac{1}{25}$.



   Tank is empty at x-intercept $(x, y) = (300, 0)$.

3. Numerical: Input-output table of the function $y = f(x)$ at sample inputs x.

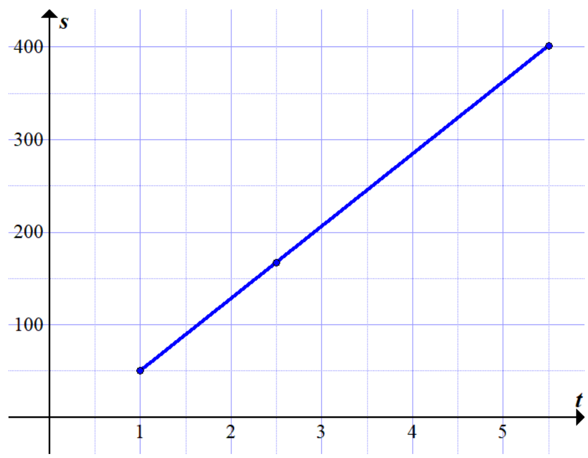
x mi	0	50	100	150	200	250	300
y gal	12	10	8	6	4	2	0

4. Algebraic: Intercept-slope formula $y = f(x) = 12 - \frac{1}{25}x$.
   Solve $y = f(x) = 0$, i.e. $12 - \frac{1}{25}x = 0$, get $x = \frac{-12}{-1/25} = 300$.

Linear function example: miles per hour

1. Physical: A car enters the highway at mile 50 at 1:00pm, and exits at mile 401 at 5:30pm. Question: What is the mile location at 2:30pm?

2. Graphical: Location is mile s at t o'clock. Linear function $s = f(t)$ between points $(t_1, s_1) = (1, 50)$ and $(t_2, s_2) = (5.5, 401)$.
   Position at $t = 2.5$ is approximately $s = f(2.5) \approx 170$.



3. Numerical: Input-output table of the function $s = f(t)$ at sample inputs t.
   speed $= \frac{\text{distance traveled}}{\text{time elapsed}} = \frac{401-50}{5.5-1.0} = \frac{351 \text{ mi}}{4.5 \text{ hr}} = 78$ mi/hr, 39 mi in 0.5 hr.

t hr	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5
s mi	50	89	128	167	206	245	284	323	362	401

   Get $s = f(2.5) = 167$.

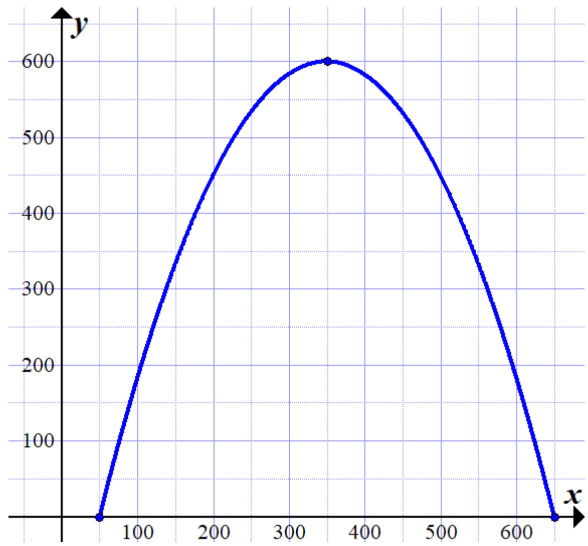
4. Algebraic: Slope between points (t_1, s_1), (t_2, s_2) is $m = \frac{s_2 - s_1}{t_2 - t_1} = 78$.
   Point-slope formula $f(t) = s_1 + m(t - t_1) = 50 + 78(t-1)$.
   Thus $f(2.5) = 50 + 78(2.5-1) = 167$.

Quadratic function example: arch

1. Physical: The Gateway Arch has the shape of a downward-cupping parabola with width and height both equal to 600 ft. One leg of the arch is positioned 50 ft from a base point, the other leg is at position 650 ft.

   Question: Find a formula $h(x)$ for the arch height above position x ft.

2. Graphical: height $y = h(x)$



3. Algebraic: We have the roots $h(x) = 0$ for $x = 50$ and $x = 650$, so the factored form is $h(x) = a(x-50)(x-650)$ for some coefficient a. The top of the arch is above the midpoint $x = \frac{50+650}{2} = 350$, so $h(350) = 600$, $a(350-50)(350-650) = 600$, $a = -\frac{1}{150}$. Thus:

$$h(x) = -\tfrac{1}{150}(x-50)(x-650) = -\tfrac{1}{150}x^2 + \tfrac{14}{3}x - \tfrac{650}{3}.$$

4. Numerical: Construction engineers use approximate $h(x)$ at sample x.

x ft	50	100	150	200	250	300	350	400	450	500	550	600	650
$h(x)$ ft	0	183	333	450	533	583	600	583	533	450	333	183	0

Rational function example: concentration of a mixture

1. Physical: A mixing tank contains a solution of 15 lb sugar in 100 gal water. The tank is then filled with 10 gal/min of water and 1 lb/min of sugar.

   Question: What is the sugar concentration after 10 min? What is the eventual concentration after a long time?

2. Numerical: Concentration function $c(t)$ lb/gal after t min. Start with $c(0) = \frac{15 \text{ lb}}{100 \text{ gal}}$, then each minute, sugar increases by 1 lb and water by 10 gal.

t min	0	1	2	3	4	5	6	7	8	9	10
$c(t)$ lb/gal	$\frac{15}{100}$	$\frac{16}{110}$	$\frac{17}{120}$	$\frac{18}{130}$	$\frac{19}{140}$	$\frac{20}{150}$	$\frac{21}{160}$	$\frac{22}{170}$	$\frac{23}{180}$	$\frac{24}{190}$	$\frac{25}{200}$
	0.150	0.145	0.141	0.138	0.135	0.133	0.131	0.129	0.127	0.126	0.125

   The concentration starts out at $c(0) = 0.15$ and decreases to $c(10) = 0.125$. Later, $c(100) = \frac{115}{1100}$, $c(1000) = \frac{1015}{10100}$, asymptotically approaching $\frac{1}{10} = 0.1$.

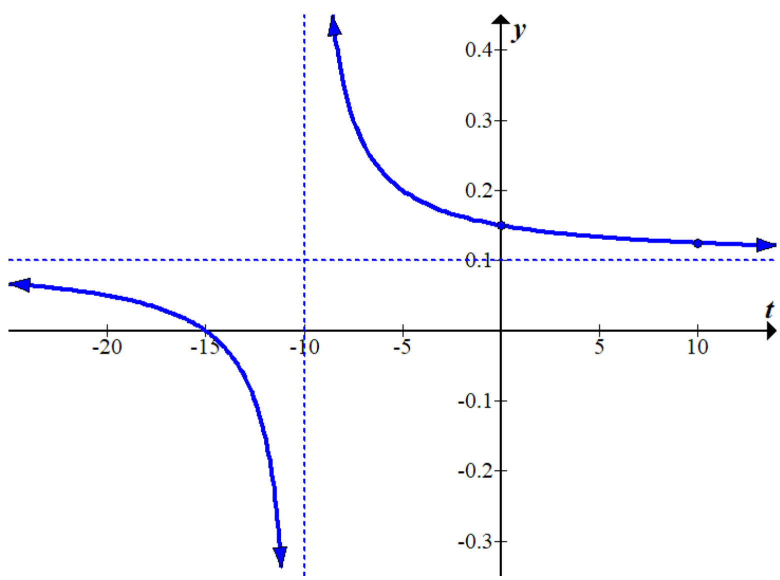
3. Algebraic: The amount of sugar is $15 + t$, the amount of water $100 + 10t$, so the concentration is $c(t) = \frac{15+t}{100+10t}$. For $t < 0$, this gives the concentration in the past, assuming the same fill rates. This makes no sense past where the denominator is zero, $t = -10$: at this time, the tank started with 5 lb sugar and no water, an infinite concentration: $c(-10) = \infty$, or more precisely, $c(t) \to \infty$ as $t \to -10$, where the arrow $\to$ means "approaches". For $t < -10$, the formula has no physical meaning.

   Polynomial long division gives $(t+15) \div (10t+100) = \frac{1}{10}$ remainder 5, so:

$$c(t) = \frac{t + 15}{10t + 100} = \frac{1}{10} + \frac{5}{10t + 100}.$$

   For very large $t \to \infty$, the second term goes to 0, so $c(t) \to \frac{1}{10}$.

4. Graphical: Plot the curve $y = c(t)$. The line where the denominator $= 0$ is a vertical asymptote $t = -10$, and the limiting value as $t \to \infty$ is a horizontal asymptote $y = \frac{1}{10}$. The graph is a hyperbola curve.



Periodic function example: hours of daylight

1. Physical: In Michigan on Jun 21, the longest day of the year, there is about $15\frac{1}{4}$ hours of daylight; and on Dec 21, the shortest day, there is about 9 hours. How long a day would we expect on Aug 1?

2. Numerical: Measuring time by days of the year (ignoring leap day), Jun 21 is day 171, and Dec 21 is day 354. Measuring daylight by hours, the maximum of the daylight function $\ell(t)$ is $\ell(171) = 15.25$; its minimum is $\ell(354) = 9$.

   Aug 1 is day 212, which is $\frac{212-171}{354-171} = 0.22$ of the way from the maximum to the minimum day, so linear interpolation suggests that $\ell(212)$ is lower than the maximum by 0.22 of the max-to-min daylight difference $15.25-9 = 6.25$:

   $$\ell(212) \;\approx\; 15.25 - \tfrac{212-171}{354-171}(15.25-9) \;=\; 15.25 - (0.22)(6.25) \;=\; 13.9 \text{ hr}\,.$$

   However, we expect this to be too low, because the daylight decreases only slowly after its Jun 21 solstice, and accelerates only near the Sep 21 equinox.

3. Algebraic. We model the undulations of daylight by a sinusoidal function:

   $$\ell(t) \;=\; a + b\sin(ct+d).$$

   We must adjust the constants a, b, c, d so that the properties of the sine function are scaled to the desired properties of $\ell(t)$.

   Sine has max/min values 1 and -1, whereas ℓ has 15.25 and 9; thus we should have $a + b = 15.25$ and $a - b = 9$, giving:

   $$a \;=\; \tfrac{15.25+9}{2} \;=\; 12.12\,, \qquad\qquad b \;=\; \tfrac{15.25-9}{2} \;=\; 3.12\,.$$

   Sine has period 2π, while ℓ has period 365, so $2\pi = 365c$. The function $\sin x$ has a maximum at $x = \frac{\pi}{2}$, while $\ell(t)$ has $t = 171$, so we want $\frac{\pi}{2} = 171c + d$:

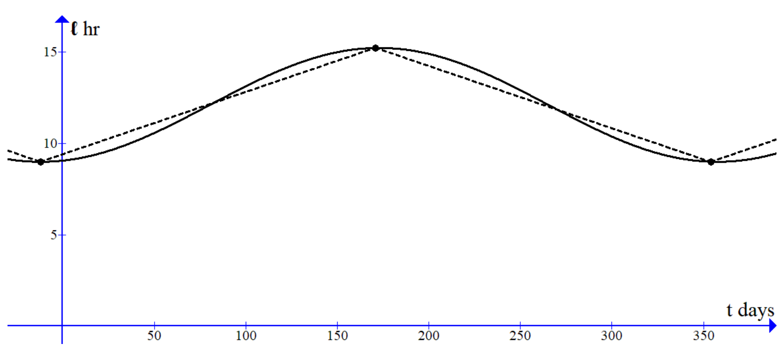
   $$c \;=\; \tfrac{2\pi}{365} \;=\; 0.017\,, \qquad d \;=\; \tfrac{\pi}{2} - 171c \;=\; \tfrac{\pi}{2} - (171)\tfrac{2\pi}{365} \;=\; -1.37\,.$$

   Therefore, our model is: $\ell(t) \;=\; 12.12 + 3.12\sin(0.017\,t-1.37)$.

   For Aug 1 = day 212, this predicts $\ell(212) = 14.6 = 14\,\text{hr}\ 36\,\text{min}$ of daylight.

   Observations give the actual time as $14.43 = 14\,\text{hr}\ 26\,\text{min}$, showing that the true daylight function is a little flatter than a sinusoid, but this is still much more accurate than the piecewise-linear model.

4. Graphical: Plot the piecewise-linear and sinusoidal model curves.



Math 132 Overview

This section is a bird's-eye view of the course. Read it over now, then come back to it as you learn the topics, to see how they fit into the whole theory.

Calculus is the mathematics of change and variation. With ordinary algebra, we can translate static or linear real-world problems into equations and solve them; with calculus, we can solve dynamic problems involving motion, rates of change, optimum values, curved shapes, and the cumulative effect of a changing influence. It was discovered by Newton and Leibnitz, and developed further notably by Euler.

The main concepts of calculus are *derivatives* and *integrals* applied to *functions*. Like most mathematical concepts, these have four levels of meaning: physical (real-world), geometric (pictures), numerical (spreadsheets), and algebraic (formulas). Given a problem originating on one level (usually physical or geometric), we translate to a different level (numerical or algebraic) where the problem can be solved, then we translate the solution back to the original level.

Functions. Officially, a function $f : X \to Y$ is any rule that takes elements of an input set X (the domain) to elements of an output set Y. In problems, this concept is represented on the following levels.

1. Physical: A function defines how an input quantity (the independent variable or argument) determines an output quantity (the dependent variable or value). For example, consider a stone dropped from a bridge: the elapsed time t (in sec) determines the observed distance s (in feet) that the stone has fallen, $s = f(t)$. If the stone falls into the water 400 ft below after 5 sec, the domain is naturally $0 \leq t \leq 5$, namely the interval $X = [0, 5]$.

2. Geometric: A function is a graph in the plane, the curve of points (x, y) such that $y = f(x)$. In our example, we use coordinates (t, s), and the graph $s = f(t)$ curves upward from $(0, 0)$ to $(5, 400)$. As the stone speeds up with increasing t, the graph gets steeper: in fact, it is a segment of a parabola.

3. Numerical: A function is a table of values. In our example, we might get a partial such table by measuring the distance at sample times:

t	0	1	2	3	4	5
$s{=}f(t)$	0	16	64	144	256	400

   Of course, $f(t)$ has a value at every t, not just the samples. We can imagine the full function as an infinite table with an entry for every t in the domain.

4. Algebraic: A function is a formula to compute the output in terms of the input. A model of our physical example is the formula $f(t) = 16t^2$. Like all models of the real world, this is accurate only within a bounded domain ($0 \leq t \leq 5$) and up to some error (from air resistance or imprecise measurements).

Notes by Peter Magyar magyar@math.msu.edu

Derivatives. Now we preview the main concepts of this course. Given a function f, the *derivative function* f' has the following meanings.

1. Physical: The derivative of a function $y = f(x)$ is the rate of change of y with respect to the change in x. In our example of a falling stone, $s = f(t)$, the derivative tells how fast the distance is increasing per unit time, i.e. how fast the stone is moving. This is the instantaneous *velocity* v in feet per second, at time t; and the derivative is the velocity function $v = f'(t)$.

2. Geometric: For a graph $y = f(x)$, the derivative $f'(a)$ at $x = a$ is the slope of the graph near $(a, f(a))$: that is, the slope of the tangent line at that point, $y = f(a) + m(x{-}a)$, where $m = f'(a)$.

3. Numerical: We can approximate the derivative $f'(a)$, instantaneous velocity, by considering an input $x = a + h$ close to a, and dividing the rise in $f(x)$ by the run in x:

$$f'(a) \approx \frac{\Delta f}{\Delta x} = \frac{f(x) - f(a)}{x - a} = \frac{f(a{+}h) - f(a)}{h} .$$

   In our example $f(t) = 16t^2$, we can compute the approximate velocity at the instant $t = 3$ sec by considering the short time interval $3 \le t \le 3.1$, and computing the distance traveled (change in distance), divided by the time elapsed:

$$v = f'(3) \cong \frac{f(3.1) - f(3)}{3.1 - 3} = \frac{153.76 - 144}{0.1} = 97.6 .$$

   That is, after falling for 3 sec, the stone is travelling about 97.6 ft/sec.

   Once we know $f'(a)$, we can use it to approximate $f(x)$ by a linear function $f(x) \approx f(a) + f'(a)(x{-}a)$ for x near a; and error sensitivity $\Delta f \approx f'(a)\,\Delta x$.

4. Algebraic: We will give methods to compute the derivative of any formula. The foundation is the precise definition: the derivative of $f(x)$ at $x = a$ is the limiting value of its rate of change over a short interval $a \le x \le a{+}h$ as the width $\Delta x = h$ becomes smaller and smaller toward zero ($h \to 0$):

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

   We use the limit definition to determine some Basic Derivatives such as $(x^p)' = px^{p-1}$, $\sin'(x) = \cos(x)$, $\cos'(x) = -\sin(x)$, and then Rules for combining them: Sum, Difference, Product $(fg)' = f'g + fg'$, Quotient $(f/g)' = (f'g - fg')/g^2$, and Chain or composition $f(g(x))' = f'(g(x))\,g'(x)$.

   For our example $f(t) = 16t^2$, we get $f'(t) = 32t$: the velocity is steadily increasing, proportional to time. This gives the exact value $f'(3) = 96$.

Integrals. Given a function g, the *integral* from $x = a$ to $x = b$ is a number denoted $\int_a^b g(x)\,dx$, and has the following meanings.

1. Physical. Suppose a quantity $z = f(x)$ is influenced linearly by another function $g(x)$ as the input goes from $x = a$ to $x = b$: i.e. each incremental change Δx leads to a small change $\Delta z \approx g(x)\,\Delta x$. Then the integral of $g(x)$ is the cumulative effect of $g(x)$, the total change in z from $x = a$ to $x = b$:

$$\int_a^b g(x)\,dx = f(b) - f(a).$$

   In our example, suppose we start by knowing the velocity of the stone, $v = g(t) = 32t$, and we wish to deduce the distance fallen, $s = f(t)$ for $t = 3$. Over a time increment Δt, the stone moves by about $\Delta s \approx v\,\Delta t = 32t\,\Delta t$; so we can express the cumulative change as: $f(3) = f(3) - f(0) = \int_0^3 32t\,dt$.

2. Geometric. For the graph $y = g(x) \geq 0$, the integral $\int_a^b g(x)\,dx$ is the area under the graph and above the interval $a \leq x \leq b$ on the x-axis. This is because the area A is the cumulative total of thin slices $\Delta A \approx g(x)\,\Delta x$ with height $y = g(x)$ and width Δx. (Area under the x-axis is counted negative.)

   In our example, we can get the integral $\int_0^3 32t\,dt$ as the triangular area under the graph $v = g(t) = 32t$ and above $t \in [0, 3]$: i.e. $\int_0^3 32t\,dt = \frac{1}{2}(3)(96)$.

3. Numerical. We approximate the cumulative effect of $g(x)$ from $x = a$ to $x = b$ by splitting up the interval $a \leq x \leq b$ into a large number n of small increments of width $\Delta x = \frac{b-a}{n}$. We take sample points $x_1, \ldots, x_n$, one in each increment, and compute the "Riemann sum" of all $\Delta z \approx g(x_i)\Delta x$:

$$\int_a^b g(x)\,dx \;\approx\; g(x_1)\Delta x + g(x_2)\Delta x + \cdots + g(x_n)\Delta x\,.$$

   This is the origin of the notation $\int_a^b g(x)\,dx$, where $\int$ is an elongated S standing for "sum," and $g(x)\,dx$ represents all the small changes $g(x_i)\Delta x$.

   In our example, given the velocity function $v = g(t) = 32t$, we can take $n = 3$, $\Delta t = 1$ sec, and sample points $t_1{=}1$, $t_1{=}2$, $t_2{=}3$. We approximate the cumulative distance traveled $\int_0^3 32t\,dt$ as the sum over the 3 time increments of: (velocity at t_i)$\times$(time elapsed) $= 32t_i\,\Delta t$, giving:

$$\int_0^3 32t\,dt \;\approx\; 32(1)(1) + 32(2)(1) + 32(3)(1) \;=\; 192.$$

   This overestimates because we sample the velocity at the end of each time increment, when the stone is fastest. Taking more increments (larger n) gives better and better approximations whose limit is the exact integral.

4. Algebraic. Since integrals go from a rate of change to a total change, they are reverse derivatives (antiderivatives), and we can use our known derivative rules backwards to find formulas for many (but not all) integrals. That is, if $g(x) = f'(x)$ for a known formula $f(x)$, then $\int_a^b g(x)\,dx = f(b) - f(a)$. This is known as the *Second Fundamental Theorem of Calculus.*

   In our example, given $v = 32t$, we can find $f(t) = 16t^2$ with $f'(t) = 32t$, so we get the exact integral value $\int_0^3 32t\,dt = f(3) - f(0) = 16(3^2) - 16(0^2) = 144$.

What does it mean?
Newton discovered it in a country garden hiding out from the plague in 1666,
consummating four thousand years of math going back
to the first Sumerian nerds who scratched out farmland measurements
and grain tallies and quadratic problems on clay tablets,
to the Egyptian priests who computed the slopes of tombs for their god king,
to Pythagoras who had a vision of numbers and shapes as the one ultimate reality,
to Euclid who built a soaring tower of theorems unshakably founded on axioms,
to the Hindus who grasped Nothing as a number, not an absence of number,
to Al-Khwarizmi the Persian who invented Qabalah and Algebra,
the Breaking and Mending Method to solve equations,
to Descartes who in the modern sunrise looked at numbers and shapes as if for
the first time and could at last explain how they describe the same deep thing.
Then Newton added that smooth functions and shapes are infinitesimally linear,
and accumulating linear increments is the inverse of taking linear rates,
insights deeper and more powerful than any before, Promethean fire
that burst open the gates to theoretical science and the Modern World.
Now if we can only learn some humility before it burns us up.

Instantaneous velocity. We start our study of the derivative with the *velocity problem:* If a particle moves along a coordinate line so that at time t, it is at position $f(t)$, then compute its velocity or speed† at a given instant.

Velocity means distance traveled, divided by time elapsed (e.g. feet per second). If the velocity changes during the time interval, then this quotient is the *average velocity.* From time $t = a$ to $t = b$, the distance traveled is the change in position $f(b) - f(a)$, and the time elapsed is $b - a$, so the average velocity is:

$$v_{\text{avg}} \;=\; \frac{f(b) - f(a)}{b - a}\,.$$

What do we mean by the instantaneous velocity at time $t = a$? We cannot compute this directly, since the particle does not move at all in an instant. Rather, we find the average velocity from $t = a$ to $t = a{+}h$, where h is a small time increment, and take the instantaneous velocity v to be the limiting value approached by the average velocities:

$$v \;=\; \lim_{h \to 0} \frac{f(a{+}h) - f(a)}{h}\,,$$

where $\lim\limits_{h \to 0}$ means "the limit as h approaches 0" of the quantity on the right.

Another way to say this is that velocity is the *rate of change* of position with respect to time: how fast the position $f(x)$ is changing per unit change in time t. Thus, v_{avg} is the average rate of change over an interval $t \in [a, b]$, while v is the instantaneous rate of change at a particular $t = a$.

Falling stone example. A stone dropped off a bridge has position approximately $f(t) = 16t^2$ feet below the bridge after falling for t seconds. The average velocity between $t = 3$ and $t = 4$ is:

$$v_{\text{avg}} \;=\; \frac{f(4) - f(3)}{4 - 3} \;=\; \frac{16(4^2) - 16(3^2)}{1} \;=\; 112.$$

That is, the stone has an average velocity of 112 ft/sec, although it starts slower than this at $t = 3$ and speeds up steadily throughout the interval.

Now, what is the instantaneous velocity at $t = 3$? We compute the average velocity over a short time interval from $t = 3$ to $t = 3 + h$, for example $h = 0.1$:

$$v_{\text{avg}} \;=\; \frac{f(3.1) - f(3)}{3.1 - 3} \;=\; \frac{16(3.1^2) - 16(3^2)}{0.1} \;=\; 97.6\;.$$

---

*Notes by Peter Magyar `magyar@math.msu.edu`, with sections corresponding to James Stewart's *Calculus*, 7th ed.

† *Velocity* can be positive or negative, depending on the direction of motion. *Speed* is the absolute value of velocity.

This is a pretty good estimate of the velocity, but to be more precise we take shorter intervals:

h	1	0.1	0.01	0.001	0.0001	0.00001
v_{avg}	112	97.6	96.16	96.016	96.0016	96.00016

It is pretty clear that as the interval gets shorter and shorter, the average velocity approaches the limiting value $v = 96$, and we define this to be the instantaneous velocity.

Let us prove this algebraically: instead of trying sample values of the time increment h, we let h be a variable:

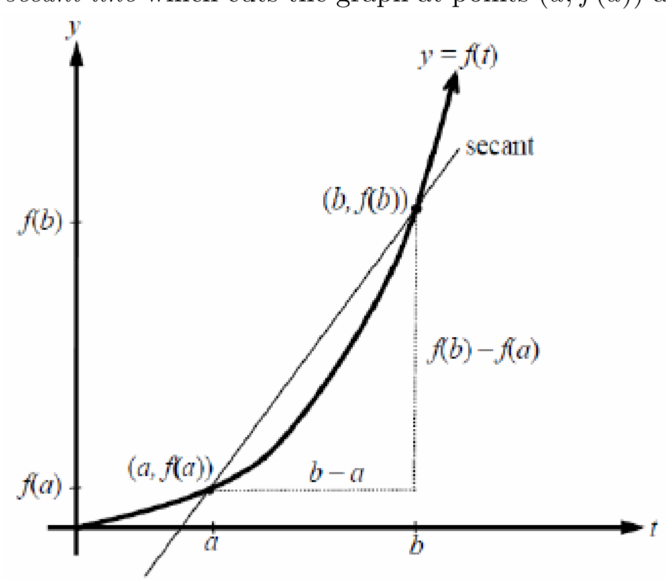
$$v_{\text{avg}} = \frac{f(3+h) - f(3)}{(3+h) - 3} = \frac{16(3+h)^2 - 16(3^2)}{h} = 16 \cdot \frac{(3+h)^2 - 3^2}{h}$$

$$= 16 \cdot \frac{(3^2 + 2(3h) + h^2) - 3^2}{h} = 16 \cdot \frac{6h + h^2}{h} = 16(6+h) = 96 + 16h\,.$$

As we take h smaller and smaller, the error term $16h$ approaches zero, and the average velocity approaches the limiting value 96, which by definition is the instantaneous velocity:

$$v = \lim_{h \to 0} \frac{f(3+h) - f(3)}{h} = 96\,.$$

Tangent Slope. We have described velocity on three conceptual levels: as a *physical* quantity, a *numerical* approximation, and an *algebraic* computation. Velocity also has a *geometric* meaning in terms of the graph $y = f(t)$. Consider a *secant line* which cuts the graph at points $(a, f(a))$ and $(b, f(b))$.

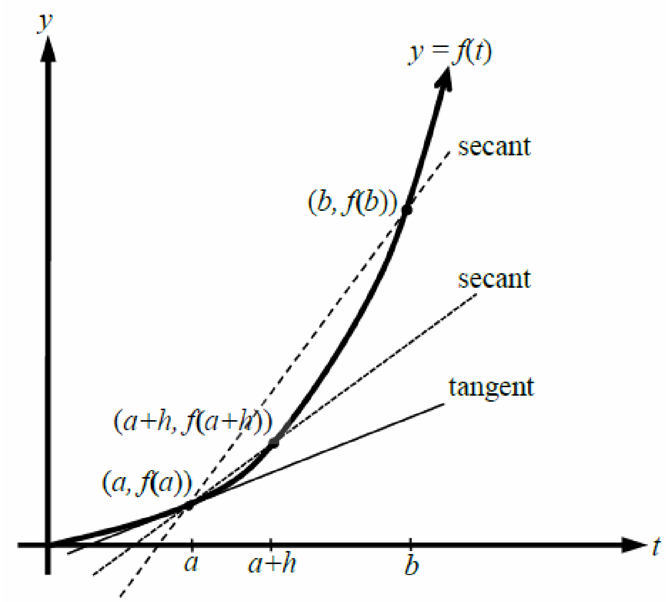


The slope m_{sec} of the secant line is the rise in the graph per unit of horizontal run, which means distance traversed divided by time elapsed, which is the average velocity:

$$m_{\text{sec}} = \frac{f(b) - f(a)}{b - a} = v_{\text{avg}}\,.$$

The reason for this coincidence is that slope is the rate of vertical rise with respect to horizontal run, just as velocity is the rate of change of position (drawn on the vertical axis) with respect to time (on the horizontal axis).

As we move the point $(b, f(b))$ to $(a{+}h, f(a{+}h))$, closer and closer to a, the secant lines approach the *tangent line* which touches the curve at the single point $(a, f(a))$.



The tangent slope m is the limit of the secant slopes, so it is equal to the instantaneous velocity:

$$m \;=\; \lim_{h\to 0} \frac{f(a{+}h) - f(a)}{h} \;=\; v\,.$$

Trig Function Example: We model average daily temperature through the year by the sinusoidal function:

$$T(x) \;=\; 55 - 40\cos(\tfrac{2\pi}{365}x) \;\; \text{degrees F on day } x.$$

How quickly is the weather warming at $x = 100$ (April 10), in degrees/day? We can only approximate the instantaneous rate of change r by looking at average rates near the given $x = 100$, for example over $x \in [100, 110]$:

$$r_1 \;=\; \frac{T(110) - T(100)}{110 - 100} \;=\; \frac{(55{-}40\cos(\frac{2\pi}{365}110)) - (55{-}40\cos(\frac{2\pi}{365}100))}{10} \;\approx\; 0.668$$

Or over $x \in [90, 100]$, giving $r_2 = 0.686$. We can get a better approximation for r by averaging the underestimate and the overestimate:

$$r \approx \tfrac{1}{2}(r_1 + r_2) = 0.677\ .$$

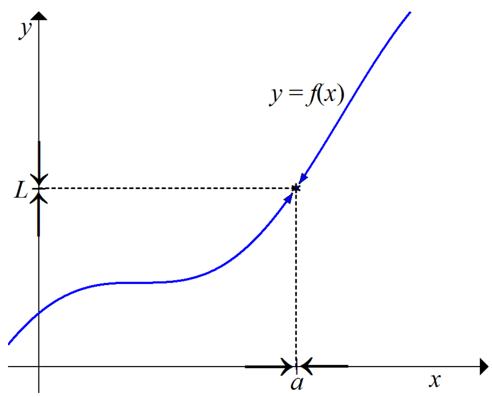
Of course, the accuracy of this model is way less than 3 decimal places, so we should say the model predicts warming of about 0.7 degree F per day.

Limits

Definition of limits. The key technical tool in the previous section was the idea of a limiting value approached by approximations. We need limits for all the definitions of calculus, so we must understand them clearly.

> *Preliminary definition:* Consider a function $f(x)$ and numbers L, a. Then the *limit* of $f(x)$ equals L as x approaches a, in symbols $\lim_{x\to a} f(x) = L$, whenever $f(x)$ can be forced arbitrarily close to L by making x sufficiently close to (but unequal to) a.

That is, $f(x)$ approximates L to within any desired error tolerance, for all values of x within some small distance from a (but $x \neq a$). One more way to say it: if we make a table of $f(x)$ for any sample values of x getting closer and closer to a (such as $x = a+0.1$, $a+0.01$, etc.), then the values of $f(x)$ will get as close as we like to L (though they might never reach L). Graphically:



Evaluating limits. Some limits are easy because we can plug in $x = a$ to get the limiting value $\lim_{x\to a} f(x) = f(a)$, in which case we say $f(x)$ is *continuous* at $x = a$. Graphically, as in the above picture, this means the curve has no jump or hole at $(a, f(a))$. For example,

$$\lim_{x\to 5} x^2 \;=\; 5^2 \;=\; 25,$$

as we could see from the graph of $y = x^2$. Algebraically, if x is close enough to 5, say $x = 5 + h$ for some small h, then

$$x^2 \;=\; (5+h)^2 \;=\; 5^2 + 2(5h) + h^2 \;=\; 25 + 10h + h^2,$$

which is forced as close as we like to $L = 25$ if h is small enough (positive or negative).

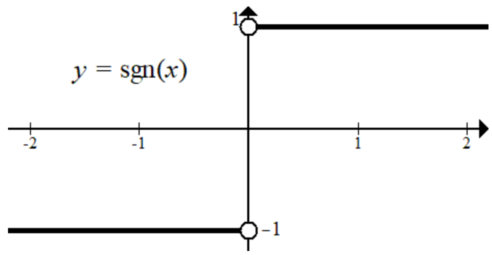
Sometimes $f(x)$ does not approach any limiting value at $x = a$, in which case we say the limit *does not exist*, and the symbol $\lim_{x\to a} f(x)$ has no

Notes by Peter Magyar `magyar@math.msu.edu`

meaning. For example, define the signum function $\mathrm{sgn}(x)$ as:

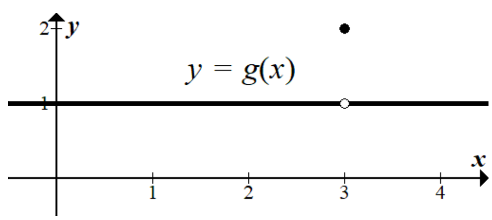
$$\mathrm{sgn}(x) \;=\; \frac{x}{|x|} \;=\; \begin{cases} +1 & \text{for } x > 0 \\ -1 & \text{for } x < 0 \\ \textit{undefined} & \text{for } x = 0, \end{cases}$$

with graph:



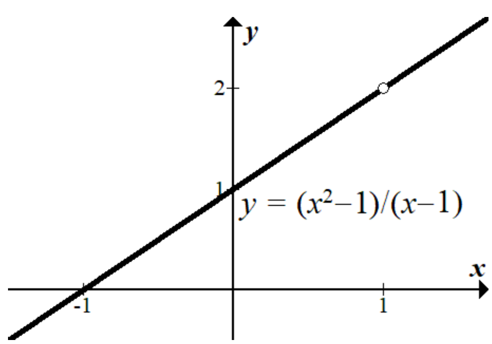
Near $x = 0$, the function cannot be forced close to any single output value. That is, $\lim_{x\to 0} \mathrm{sgn}(x) \neq 1$, since no matter how close x gets to 0, there are some x (namely negative) for which $\mathrm{sgn}(x)$ is far from 1; and similarly $\lim_{x\to 0} \mathrm{sgn}(x)$ is not -1, nor 0, nor any other value. In particular, it is *false* that $\lim_{x\to 0} \mathrm{sgn}(x) = \mathrm{sgn}(0)$, and the function is not continuous at $x = 0$.

An important feature of $\lim_{x\to a} f(x)$ is that it does not depend on $f(a)$, even if $f(a)$ is undefined: the limit only notices values of $f(x)$ for $x \neq a$. For example, define $g(x) = 1$ for $x \neq 3$, and $g(3) = 2$, having the graph:



Then $\lim_{x\to 3} g(x) = 1$, since if x is close enough to (but unequal to) 3, then $g(x)$ is arbitrarily close to $L = 1$ (in fact $g(x) = L$). Again, $\lim_{x\to 3} g(x) \neq g(3) = 2$, and $g(x)$ is not continuous at $x = 3$.

The important limits in calculus, such as instantaneous velocity, are cases where the function is not defined at $x = a$. For example, consider $\lim_{x\to 1} \frac{x^2-1}{x-1}$. Plugging in $x = 1$ gives the meaningless expression $\frac{0}{0}$, so this function is not continuous, but the limit still exists. Indeed, plotting points gives the graph:



It seems the limit is $L = 2$: the graph approaches $(1, 2)$, so if x is sufficiently close to (but not equal to) 1, then $f(x)$ is forced as close as desired to 2. We can prove this algebraically:

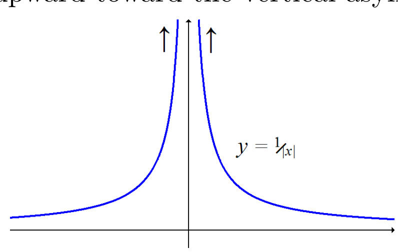
$$\lim_{x\to 1} \frac{x^2-1}{x-1} \;=\; \lim_{x\to 1} \frac{(x-1)(x+1)}{x-1} \;=\; \lim_{x\to 1} x+1 \;=\; 1+1 \;=\; 2,$$

since $x+1$ is continuous.

One-sided and infinite limits. We define another type of limit. *One-sided limits* (from the right or left) notice only values of x on one side of a. That is, the limit of $f(x)$ equals L as x approaches a from the right, denoted $\lim_{x\to a^+} f(x) = L$, whenever $f(x)$ can be forced arbitrarily close to L by making x sufficiently close to (but *greater* than) a. The limit from the left, denoted $\lim_{x\to a^-} f(x) = L$, is the same, except with x *less* than a.

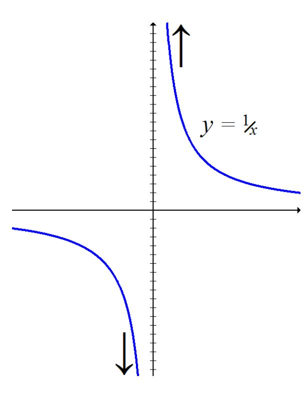
If we have the ordinary limit $\lim_{x\to a} f(x) = L$, then clearly the left and right limits have the same value L. Thus, in the above examples, we have $\lim_{x\to 5^+} x^2 = \lim_{x\to 5^-} x^2 = 5^2$, and $\lim_{x\to 1^+} g(x) = \lim_{x\to 1^-} g(x) = 0$, and $\lim_{x\to 1^+} \frac{x^2-1}{x-1} = \lim_{x\to 1^-} \frac{x^2-1}{x-1} = 2$. However, $\lim_{x\to 0^+} \mathrm{sgn}(x) = 1$ and $\lim_{x\to 0^-} \mathrm{sgn}(x) = -1$, even though $\lim_{x\to 0} \mathrm{sgn}(x)$ does not exist.

Finally, we define *infinite limits*: $\lim_{x\to a} f(x) = \infty$ means that $f(x)$ can be forced larger than any bound (for instance $f(x) > 1000$) by making x sufficiently close to (but not equal to) a. The symbol ∞ has no meaning by itself: this is just a way of saying that $f(x)$ becomes as large a number as we like. For example, we have $\lim_{x\to 0} \frac{1}{|x|} = \infty$, since $\frac{1}{\text{tiny}} = \text{huge}$, so the graph $y = \frac{1}{|x|}$ shoots upward toward the vertical asymptote $x = 0$.



However, for the function $\frac{1}{x}$, we have $\lim_{x\to 0} \frac{1}{x} \neq \infty$, since no matter how close x is to 0, we cannot force $\frac{1}{x}$ above a given positive bound: rather, for x a tiny negative number, $\frac{1}{x} = \frac{1}{-\text{tiny}} = -\text{huge}$, a large *negative* number. In fact, the graph shoots upward to the right of the vertical asymptote, and downward to the left of the asymptote, so we have one-sided infinite limits:

$$\lim_{x\to 0^-} \frac{1}{x} \;=\; -\infty$$



$$\lim_{x\to 0^+} \frac{1}{x} \;=\; \infty$$

Vertical asymptotes. We determine the asymptotic behavior of:

$$f(x) \;=\; \frac{2x-4}{x^2-4x+3} \;=\; \frac{2(x{-}2)}{(x{-}1)(x{-}3)}\,.$$

Given the first form of the function, we immediately factor to see the vanishing of the numerator at $x = 2$ and the denominator at $x = 1, 3$. (These x-values are different, so no factors cancel.) The vanishing of the numerator shows when $f(x) = 0$, namely at the x-intercept $x = 2$.

The vanishing of the denominator shows when $f(x)$ becomes huge, namely near the vertical asymptotes $x = 1$ and $x = 3$. To see whether the function goes up or down near the asymptotes, we keep track of the signs.

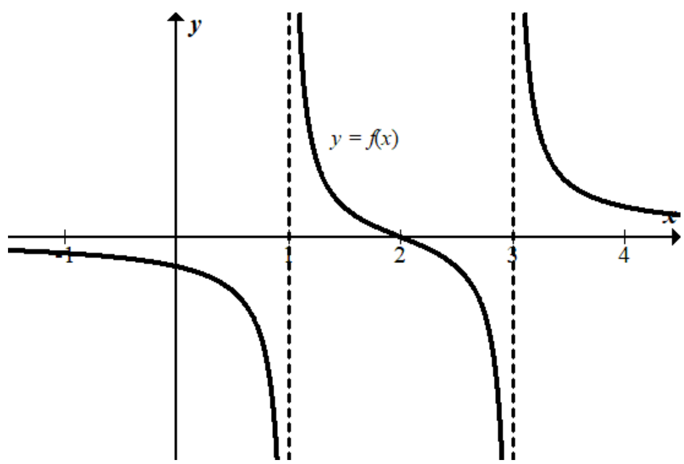
For $x < 1$, we have $x{-}1$, $x{-}2$, $x{-}3 < 0$ all negative:

$$f(x) \;=\; \frac{2(x{-}2)}{(x{-}1)(x{-}3)} \;=\; \frac{2(-)}{(-)(-)} \;=\; (-) \quad \text{so} \quad \lim_{x\to 1^-} f(x) = -\infty.$$

For $1 < x < 2$, we have $x{-}1 > 0$ and $x{-}2$, $x{-}3 < 0$:

$$f(x) \;=\; \frac{2(x{-}2)}{(x{-}1)(x{-}3)} \;=\; \frac{2(-)}{(+)(-)} \;=\; (+) \quad \text{so} \quad \lim_{x\to 1^+} f(x) = \infty.$$

Similarly, $\lim_{x\to 3^-} f(x) = -\infty$ and $\lim_{x\to 3^+} f(x) = \infty$. The graph is:



Of course, as for all limits we can approximate by plugging in sample inputs: for example, $f(.9) \approx -10.5$, $f(.99) \approx -100.5$, so it seems $\lim_{x\to 1^-} f(x) = -\infty$.

NOTE: In the slightly different function $\frac{2x-2}{x^2-4x+3} = \frac{2(x-1)}{(x-1)(x-3)}$, both numerator and denominator vanish at $x = 1$. The $(x{-}1)$ factors cancel, and the function has neither an asymptote nor an intercept at $x = 1$, only a hole in the graph where $f(1) = \frac{0}{0}$ is undefined.

Operations on limits. Some general combination rules make most limit computations routine. Suppose we know that $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist. Then we have the Limit Laws:

- *Sum:* $\lim_{x\to a}(f(x)+g(x)) \;=\; \lim_{x\to a} f(x) \,+\, \lim_{x\to a} g(x)$.
- *Difference:* $\lim_{x\to a}(f(x)-g(x)) \;=\; \lim_{x\to a} f(x) \,-\, \lim_{x\to a} g(x)$.
- *Constant Multiple:* $\lim_{x\to a}(c\,f(x)) \;=\; c\,\lim_{x\to a} f(x)$, for a constant c.
- *Product:* $\lim_{x\to a} f(x)g(x) \;=\; \lim_{x\to a} f(x)\cdot \lim_{x\to a} g(x)$.
- *Quotient:* $\lim_{x\to a} \frac{f(x)}{g(x)} \;=\; \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}$, provided $\lim_{x\to a} g(x) \neq 0$.
- *Power:* $\lim_{x\to a} f(x)^n \;=\; (\lim_{x\to a} f(x))^n$, for a whole number n.
- *Root:* $\lim_{x\to a} \sqrt[n]{f(x)} \;=\; \sqrt[n]{\lim_{x\to a} f(x)}$, for a whole number* n.

These all have the form: "The limit of an operation equals the operation applied to the limits." These Laws are also valid for one-sided limits.

Limits by plugging in. Assuming the Limit Laws and the Basic Limits $\lim_{x\to a} x = a$ and $\lim_{x\to a} c = c$, we can prove that most functions are continuous, meaning the $\lim_{x\to a} f(x)$ is obtained by substituting $x = a$ to get $f(a)$. For example, we can formally compute the limit:

$$\lim_{x\to 2} \frac{1-\sqrt{x}}{1+x} = \frac{\lim\limits_{x\to 2} 1-\sqrt{x}}{\lim\limits_{x\to 2} 1+x} \qquad \text{by the Quotient Law}^{\dagger}$$

$$= \frac{\lim\limits_{x\to 2} 1 - \lim\limits_{x\to 2} \sqrt{x}}{\lim\limits_{x\to 2} 1 + \lim\limits_{x\to 2} x} \qquad \text{by the Sum and Difference Laws}$$

$$= \frac{\lim\limits_{x\to 2} 1 - \sqrt{\lim\limits_{x\to 2} x}}{\lim\limits_{x\to 2} 1 + \lim\limits_{x\to 2} x} \qquad \text{by the Root Law}$$

$$= \frac{1-\sqrt{2}}{1+2} = \frac{1-\sqrt{2}}{3} \qquad \text{by the Basic Limits.}$$

Notes by Peter Magyar magyar@math.msu.edu

*If n is even, we assume $\lim_{x\to a} f(x) > 0$.

†The Quotient Law requires that the denominator have a non-zero limit. We tentatively proceed with the computation and find the denominator to be 3, which retrospectively justifies the quotient step.

That is, the correct limit would be obtained just by substituting $x = 2$. In general, substituting $x = a$ gives the correct limit unless it leads to a meaningless expression like $\frac{0}{0}$ or $\sqrt{-1}$ (we do not consider imaginary numbers in this course). In Notes §2.4, we will show that trigonometric functions like $\sin(x)$ and $\tan(x)$ are also continuous when defined, and the same for functions like 2^x and $\log(x)$, so this principle works for pretty much all formulas.

Limits by canceling zeroes. As we have seen, the most important limits are those for which substitution gives the meaningless expression $\frac{0}{0}$. To compute these, we must cancel vanishing factors from the top and bottom, until we get an expression which can be evaluated by the Laws. This often requires factoring, for example:

$$\lim_{x\to 2} \frac{x^2-4x+4}{x^2-x-2} \;=\; \lim_{x\to 2} \frac{(x{-}2)^2}{(x{-}2)(x{+}1)} \;=\; \lim_{x\to 2} \frac{x{-}2}{x{+}1}\,,$$

which can be evaluated by substituting $x = 2$. Another trick to avoid $\frac{0}{0}$ is to multiply by a conjugate radical to eliminate square roots:

$$\lim_{x\to 9} \frac{x{-}9}{\sqrt{x}{-}3} \;=\; \lim_{x\to 9} \frac{x{-}9}{\sqrt{x}{-}3}\cdot\frac{\sqrt{x}{+}3}{\sqrt{x}{+}3} \;=\; \lim_{x\to 9} \frac{(x{-}9)(\sqrt{x}{+}3)}{(\sqrt{x})^2-3^2}$$

$$=\; \lim_{x\to 9} \frac{(x{-}9)(\sqrt{x}{+}3)}{x-9} \;=\; \lim_{x\to 9} \sqrt{x}+3 \;=\; \sqrt{9}+3 \;=\; 6.$$

Limits by cases. A familiar function defined by cases is the absolute value: $|x| = x$ for $x \geq 0$, and $|x| = -x$ for $x < 0$. To evaluate limits involving such functions, we must consider these cases separately. For example, compute:

$$\lim_{x\to 2} \frac{|x^2{-}4|}{x-2}\,.$$

The function is not continuous: plugging in $x = 2$ gives $\frac{0}{0}$. Rather, we must examine the cases where $x < 2$ and $x > 2$. In the first case, we deduce:†

$$x < 2 \;\implies\; x^2 < 4 \implies\; x^2{-}4 < 0 \;\implies\; |x^2{-}4| = -(x^2{-}4).$$

Thus, we have the left limit:

$$\lim_{x\to 2^-} \frac{|x^2{-}4|}{x-2} \;=\; \lim_{x\to 2^-} \frac{-(x^2{-}4)}{x-2} \;=\; \lim_{x\to 2^-} \frac{-(x{-}2)(x{+}2)}{x-2} \;=\; \lim_{x\to 2^-} -(x{+}2) \;=\; -4.$$

We can check this by plugging in values like $x = 1.9, 1.99, \ldots$, getting $\frac{|x^2-4|}{x-2} = -3.9,\ -3.99, \ldots \to -4$.

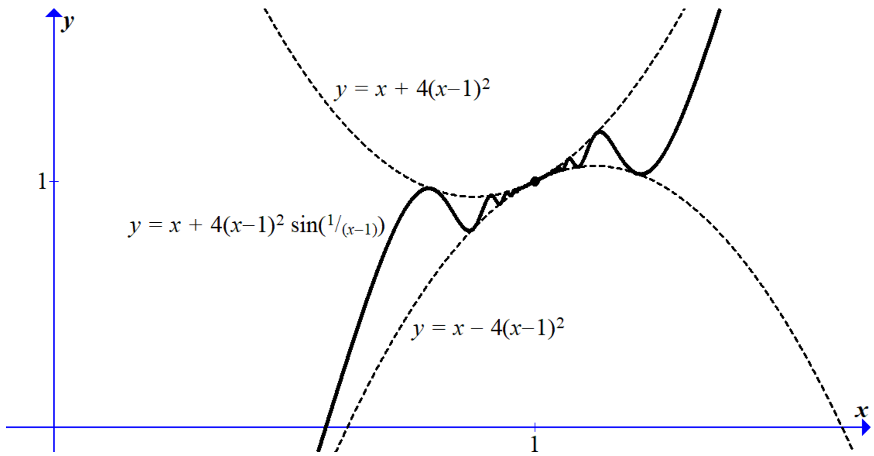
Reasoning similarly, we find $\lim_{x\to 2^+} \frac{|x^2-4|}{x-2} = \lim_{x\to 2^+}(x{+}2) = 4$. Since the one-sided limits disagree, the two-sided limit does not exist.

†The symbol $A \implies B$ means "statement A logically implies statement B".

Limits by squeezing. Some limits $\lim_{x\to a} f(x)$ are difficult to evaluate because $f(x)$ behaves erratically near $x = a$. For example:

$$\lim_{x\to 1} x + 4(x-1)^2 \sin(\tfrac{1}{x-1}).$$

The graph shows the weirdness, oscillating faster and faster near $x = 1$ because $\sin(\theta)$ goes through infinitely many periods as $\theta = \frac{1}{x-1}$ becomes larger.



But however complicated this behavior, we know $-1 \leq \sin(\theta) \leq 1$ for any θ, so our function has simple lower and upper bounds (floor and ceiling):

$$x - 4(x-1)^2 \;\leq\; x + 4(x-1)^2 \sin(\tfrac{1}{x-1}) \;\leq\; x + 4(x-1)^2.$$

Since our function lies between these bounds, so does its limit, if it exists:

$$1 \;=\; \lim_{x\to 1} x - 4(x-1)^2 \;\leq\; \lim_{x\to 1} x + 4(x-1)^2 \sin(\tfrac{1}{x-1}) \;\leq\; \lim_{x\to 1} x + 4(x-1)^2 = 1.$$

But the floor and ceiling both approach the limit $L = 1$, so our function is squeezed toward this same value:

$$\lim_{x\to 1} x + 4(x-1)^2 \sin(\tfrac{1}{x-1}) \;=\; 1.$$

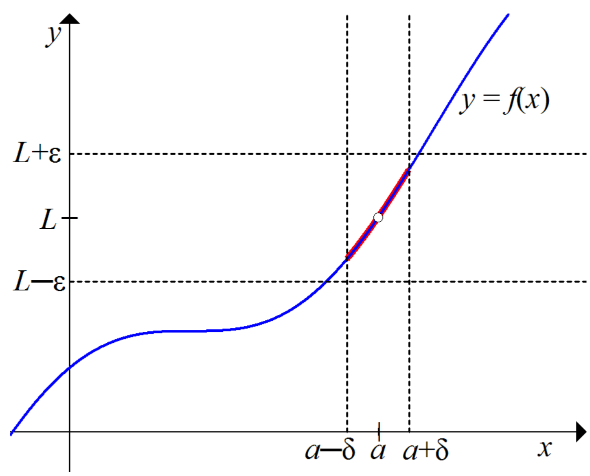
This reasoning is formalized in the following theorem.

> *Squeeze Theorem:* Suppose $g(x) \leq f(x) \leq h(x)$ for all x near a (except possibly $x = a$), and $\lim_{x\to a} g(x) = \lim_{x\to a} h(x) = L$. Then $\lim_{x\to a} f(x) = L$.

Why do we need limits? Because we cannot directly evaluate important quantities like instantaneous velocity or tangent slope, but we can approximate them with arbitrary accuracy. A limit pinpoints the exact value within this cloud of approximations. In this section, we get to the logical core of this concept.

For example, consider the tangent line of $y = x^2$ at $x = 1$, approximated by the secant through $(1, 1)$ and a nearby point (x, x^2), giving the slope: $f(x) = \frac{x^2-1}{x-1}$. There is no defined value for $f(1)$, but as x gets very close to 1, we expect the approximations $f(x)$ to have the exact tangent slope as their "limiting value", $\lim_{x\to 1} f(x) = L$. This means a candidate value L is the correct value if we can force $f(x)$ as close as desired to L (within an error $\varepsilon = \frac{1}{10}$, or $\frac{1}{100}$, or $\frac{1}{1{,}000{,}000}$, or any possible $\varepsilon > 0$), provided we restrict x close enough to 1.

Thus, proving a limit is an *error-control problem* of a type we see in the real world. For example, how accurately must you set the angle of your tennis racket to land the ball within one foot of a given spot (or within one inch)? In the general situation, an input setting x produces an output $f(x)$: how accurate must the input be to ensure a tolerable output error? That is, what allowed difference δ of x from a will force an error less than ε of $f(x)$ from L?*



In the graph $y = f(x)$, we take the small red piece between the vertical lines $a - \delta < x < a + \delta$ (not including $x = a$). By setting δ small enough, we try to force this piece between the fixed horizontal lines $L - \varepsilon < y < L + \varepsilon$, for the specified output error ε.

Rewriting $a-\delta < x < a+\delta$ as $|x - a| < \delta$, and $L-\varepsilon < f(x) < L+\varepsilon$ as $|f(x) - L| < \varepsilon$, we get the formal definition of a limit:

Notes by Peter Magyar `magyar@math.msu.edu`

*Here δ (delta) is a Greek letter d, standing for "difference", and ε (epsilon) is Greek e, standing for "error".

Definition: $\lim_{x\to a} f(x) = L$ means that for any output error tolerance $\varepsilon > 0$, there is an input accuracy $\delta > 0$ such that $0 < |x - a| < \delta$ forces $|f(x) - L| < \varepsilon$.

We can define one-sided and infinte limits similarly.[†]

Proof of Individual Limits. The precise definition allows us to rigorously prove facts about limits: specific limit computatons, as well as the general Limit Laws, which can then be applied instead of case-by-case proofs.

EXAMPLE: We prove that $\lim_{x\to 5}(3x{-}2) = 3(5){-}2 = 13$. We treat the desired error tolerance ε as a variable, and we want to guarantee the output error $|f(x) - L| < \varepsilon$, or equivalently $-\varepsilon < f(x) - L < \varepsilon$. We write this out and solve the inequalities for x:

$$-\varepsilon < (3x{-}2) - 13 < \varepsilon \iff 15 - \varepsilon < 3x < 15 + \varepsilon$$

$$\iff \tfrac{1}{3}(15{-}\varepsilon) < x < \tfrac{1}{3}(15{+}\varepsilon) \iff 5{-}\tfrac{1}{3}\varepsilon < x < 5{+}\tfrac{1}{3}\varepsilon.$$

(Here $\iff$ means "is logically equivalent to".) Finally, we put this in terms of the input accuracy $x - a = x - 5$:

$$-\tfrac{1}{3}\varepsilon < x - 5 < \tfrac{1}{3}\varepsilon.$$

To force this, we are allowed to set any input accuracy $|x - a| < \delta$, or $-\delta < x - a < \delta$. Evidently, $\delta = \frac{1}{3}\varepsilon$ will work.[‡]

EXAMPLE: A harder error-control problem: $\lim_{x\to 3}\sqrt{x} = \sqrt{3}$. We translate the output accuracy requirement $-\varepsilon < f(x) - L < \varepsilon$ into inequalities bounding the input accuracy $x - a$. (Here x is a positive value close to 3, and we take any small error tolerance $1 > \varepsilon > 0$.)

$$\begin{aligned}
-\varepsilon \;<\; \sqrt{x} - \sqrt{3} \;<\; \varepsilon \quad &\iff \quad \sqrt{3} - \varepsilon \;<\; \sqrt{x} \;<\; \sqrt{3} + \varepsilon \\
&\iff \quad \sqrt{3}^2 - 2\varepsilon\sqrt{3} + \varepsilon^2 \;<\; x \;<\; \sqrt{3}^2 + 2\varepsilon\sqrt{3} + \varepsilon^2 \\
&\iff \quad -2\varepsilon\sqrt{3} + \varepsilon^2 \;<\; x{-}3 \;<\; 2\varepsilon\sqrt{3} + \varepsilon^2
\end{aligned}$$

We need an input accuracy δ which guarantees the last inequalities above: In general, to guarantee a desired equality of the form $-d_1 < x{-}a < d_2$, we

[†] $\lim_{x\to a^+} f(x) = L$ means that for any $\varepsilon > 0$, there is some $\delta > 0$ such that $0 < x{-}a < \delta$ implies $|f(x) - L| < \varepsilon$; and $\lim_{x\to a} f(x) = \infty$ means that for any bound B, there is some $\delta > 0$ such that $0 < |x - a| < \delta$ implies $f(x) > B$.

[‡] For the formal proof, we must reverse this logic, and show that the given input accuracy guarantees the desired output accuracy. Given any desired $\varepsilon > 0$, we define $\delta = \frac{1}{3}\varepsilon$, and assume $|x - 5| < \delta$. Then we have:

$$3|x - 5| < 3\delta = \varepsilon \implies |3x - 15| < \varepsilon \implies |(3x{-}2) - 13| < \varepsilon,$$

which is our desired conclusion $|f(x) - L| < \varepsilon$. (Here $\implies$ means "logically implies".)

choose δ to be the smaller of d_1 and d_2. Thus we take $\delta = 2\varepsilon\sqrt{3} - \varepsilon^2$. Then $-\delta < x{-}3$ is equivalent to the desired lower bound $-2\varepsilon\sqrt{3} + \varepsilon^2 < x{-}3$; and also $x{-}3 < \delta$ implies the desired upper bound, since:

$$\delta \;=\; 2\varepsilon\sqrt{3} - \varepsilon^2 \;<\; 2\varepsilon\sqrt{3} + \varepsilon^2.$$

Note: In evaluating limits, we almost always rely on the Limit Laws and other general theorems, without a specific error analysis. The general results guarantee that the error approaches zero, and this is all we need.

Proof of Limit Theorems. All the general Limit Laws of §1.6 can be rigorously proved by error-control analysis. We prove the simplest one:

> *Sum Law:* If $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$, then:
>
> $$\lim_{x\to a}(f(x) + g(x)) = L + M.$$

Proof. Consider any $\varepsilon > 0$. Since we assume $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$, we can require the error tolerance $\frac{1}{2}\varepsilon$ for these limits, getting $\delta > 0$ small enough that $0 < |x - a| < \delta$ forces:

$$-\tfrac{1}{2}\varepsilon < f(x){-}L < \tfrac{1}{2}\varepsilon \qquad \text{and} \qquad -\tfrac{1}{2}\varepsilon < g(x){-}M < \tfrac{1}{2}\varepsilon.$$

Adding these inequalities, we find that $0 < |x - a| < \delta$ also forces:

$$-\tfrac{1}{2}\varepsilon - \tfrac{1}{2}\varepsilon \;<\; (f(x){-}L) + (g(x){-}M) \;<\; \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon$$

Rewriting, this is just $-\varepsilon \;<\; (f(x){+}g(x)) - (L{+}M) \;<\; \varepsilon$, which is the desired output error bound.

> *Squeeze Theorem:* If $f(x) < g(x) < h(x)$ for all values of x near a (except perhaps $x = a$), and $\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = L$, then $\lim_{x\to a} g(x) = L$.

Proof. Consider any $\varepsilon > 0$. Since we assume $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} h(x) = L$, we can find a $\delta > 0$ such that $0 < |x{-}a| < \delta$ forces $-\varepsilon < f(x){-}L < \varepsilon$ and $-\varepsilon < g(x){-}L < \varepsilon$. We also know $f(x) < g(x) < h(x)$ provided $|x - a| < \delta$ restricts x close enough to a, so:

$$f(x){-}L < g(x){-}L < h(x){-}L.$$

Then $0 < |x - a| < \delta$ also forces:

$$-\varepsilon < f(x){-}L < g(x){-}L \qquad \text{and} \qquad g(x){-}L < h(x){-}L < \varepsilon,$$

which gives the desired output accuracy for $g(x)$.

Substitution Theorem: If $\lim_{x\to b} f(x) = L$ and $\lim_{x\to a} g(x) = b$, and $g(x) \neq b$ for all x close enough to (but unequal to) a, then $\lim_{x\to a} f(g(x)) = L$.

Proof. For any $\varepsilon > 0$, we must find a number $\delta > 0$ such that $0 < |x-a| < \delta$ forces $|f(g(x)) - L| < \varepsilon$.

Take any $\varepsilon > 0$. Since $\lim_{x\to b} f(x) = L$, there is $\delta_1 > 0$ such that $0 < |y - b| < \delta_1$ forces $|f(y) - L| < \varepsilon$. Also, since $\lim_{x\to a} g(x) = b$, there exists $\delta_2 > 0$ such that $0 < |x - a| < \delta_2$ forces $|g(x) - b| < \delta_1$. Now take $\delta < \delta_2$, and δ small enough that $0 < |x - a| < \delta$ forces $g(x) \neq b$. Then we know $|x - a| < \delta$ forces $0 < |g(x) - b| < \delta_1$, which in turn forces $|f(g(x)) - L| < \varepsilon$, as required.

Math 132

Continuity

Stewart §1.8

One of the most basic features of a function is whether it is *continuous*. Roughly, this means that a small change in x always leads to a fairly small change in $f(x)$, without instantaneous jumps. In physical terms, the position of a particle moving in space is continuous, but the position displayed in a video could have a gap, making the position function jump discontinuously. This can be made precise by saying that near $x = a$, the limit of $f(x)$ is $f(a)$:

Definition: A function $f(x)$ is continuous at $x = a$ whenever $\lim_{x\to a} f(x) = f(a)$.

Graphically, a function is continuous whenever the graph $y = f(x)$ proceeds through the point $(a, f(a))$ without jumps or holes.

Types of discontinuity. If $f(x)$ is defined near $x = a$, continuity can fail in several ways:

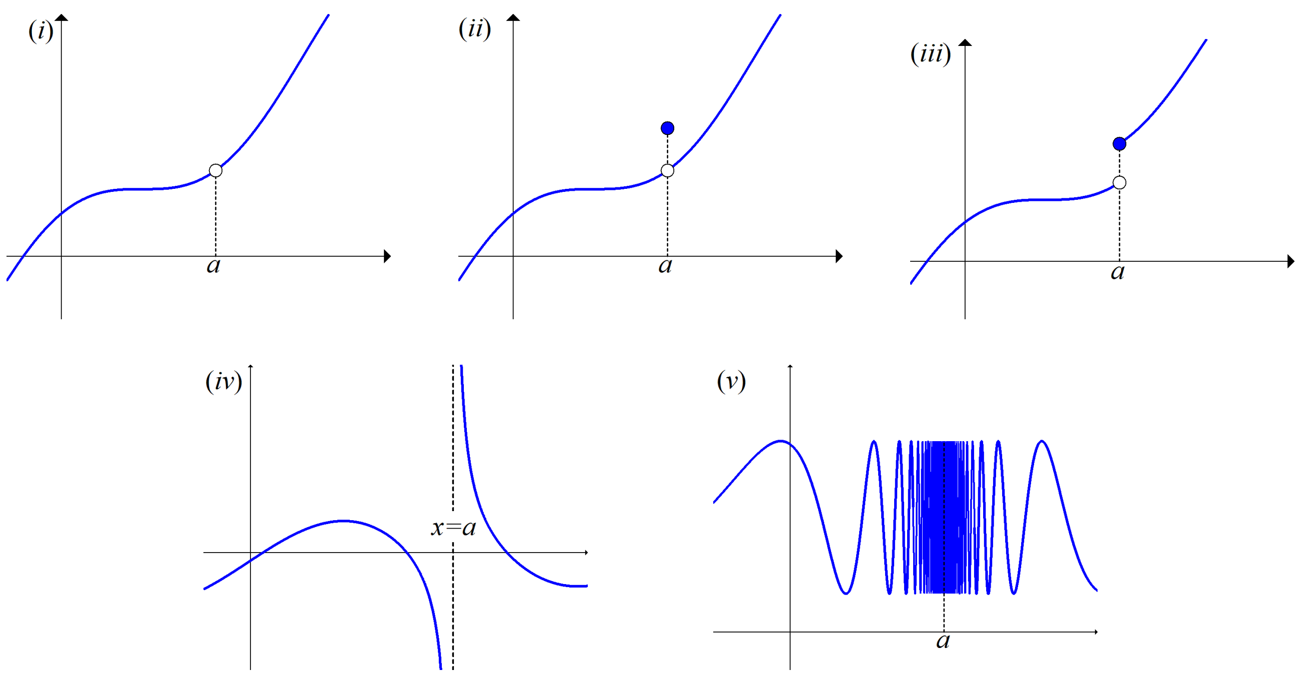
i. Removable discontinuity: $f(a)$ is undefined, but $\lim_{x\to a} f(x)$ exists.

ii. Removable discontinuity: $f(a)$ and $\lim_{x\to a} f(x)$ exist, but are unequal.

iii. Jump discontinuity: the left and right limits are unequal, $\lim_{x\to a^+} f(x) \neq \lim_{x\to a^-} f(x)$.

iv. Vertical asymptote: $\lim_{x\to a^+} f(x)$ and/or $\lim_{x\to a^-} f(x)$ are $\pm\infty$.

v. Essential discontinuity: $\lim_{x\to a^+} f(x)$ and/or $\lim_{x\to a^-} f(x)$ do not exist.



We say $f(x)$ is *continuous on an interval* whenever it is continuous at each point of the interval. For the endpoints of a closed interval* $x \in [a, b]$, we cannot take two-sided limits within the interval, so we only require $\lim_{x\to a^+} f(x) = f(a)$ and $\lim_{x\to b^-} f(x) = f(b)$.

Notes by Peter Magyar `magyar@math.msu.edu`

*The interval $[a, b]$ is the *set* or collection of all numbers x between a and b, including the endpoints. The notation $x \in [a, b]$ means x is an *element* of the set $[a, b]$, meaning it is one of the numbers between a and b, which means $a \le x \le b$.

Continuity by cases. For a function defined by cases, whose graph consists of several continuous pieces, the function is continuous provided the pieces join together at the transition points where cases meet.

For example, suppose a weight is winched from the ground at a constant speed for 8 seconds, is dropped, and lands 2 sec later. How fast should the winch haul upward?

The height $s(t)$ feet at t sec is given by:

$$s(t) \;=\; \begin{cases} vt & \text{for } t \in [0,8) \\ 64 - 16(t{-}8)^2 & \text{for } t \in [8,10]. \end{cases}$$

Here the first case is the lift at constant velocity v. For the second case, recall that a dropped weight falls $16t^2$ ft in t sec, so the 2 sec fall must start at height $16(2^2) = 64$ ft, then drop by $16(t{-}8)^2$ ft at each time past $t = 8$.

We seek the correct speed v to make the rising and falling pieces join at $t = 8$: that is, $v \cdot 8 = s(8) = 64 - 16(8{-}8)^2$, so $v = \frac{64}{8} = 8$ ft/sec. This is continuous at $t = 8$ since: $\lim_{t\to 8^-} s(t) = \lim_{t\to 8^-} 8t = 64$, and $\lim_{t\to 8^+} s(t) = \lim_{t\to 8^+} 64 - 16(t{-}8)^2 = 64$, so the two-sided limit is $\lim_{t\to 8} s(t) = 64 = s(8)$.

Domain of continuity. Almost all functions defined by formulas are continuous, except at points where they are undefined. This follows from our methods for computing limits.

EXAMPLE: Find the points where the following function is continuous:

$$g(x) = \frac{(x^2{-}3x{+}1)\sqrt{x{+}1}}{x-3}\,.$$

First, we consider the factors outside the square root, repeatedly applying the Limit Laws from §1.6:

$$\lim_{x\to a} \frac{x^2 - 3x + 1}{x - 3} \;=\; \frac{(\lim_{x\to a} x)^2 - 3\,(\lim_{x\to a} x) + 1}{(\lim_{x\to a} x) - 3} \;=\; \frac{a^2 - 3a + 1}{a - 3},$$

provided the denominator $a{-}3$ is non-zero; that is, $a \neq 3$. The Limit Laws also give $\lim_{x\to a} \sqrt{x{+}1} = \sqrt{a{+}1}$ provided $a{+}1 > 0$ to avoid the square root of a negative number; that is, for $a > -1$. Combining these, we have:

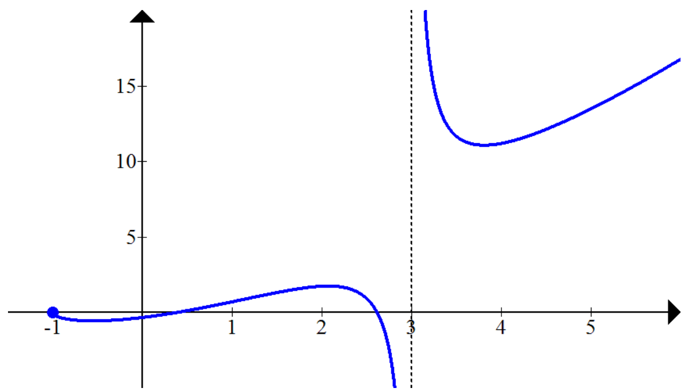
$$\lim_{x\to a} \frac{(x^2{-}3x{+}1)\sqrt{x{+}1}}{x-3} \;=\; \lim_{x\to a} \frac{x^2 - 3x + 1}{x - 3} \cdot \lim_{x\to a} \sqrt{x{+}1} \;=\; \frac{(a^2{-}3a{+}1)\sqrt{a{+}1}}{a-3}\,,$$

provided both factor limits exist, that is if $a \neq 3$ and $a > -1$. That is, $g(x)$ is continuous for all these values of a. The remaining values are:

- $a < -1$, where $g(x)$ is undefined, hence not continuous;
- $a = -1$, where $g(x)$ is continuous, since at the left endpoint of the domain of definition, we only require the one-sided limit $\lim_{x\to a^+} g(x) = g(a)$;
- $a = 3$, where the function clearly has a vertical asymptote, discontinuity of type (iv).

In summary, our $g(x)$ is continuous at every point where it is defined, that is, in the intervals† $[-1, 3) \cup (3, \infty)$. The graph looks like:

†The half-open interval $[a, b)$ is the set of all numbers x between a and b, including the left endpoint $x = a$ but excluding the right endpoint $x = b$; that is, $a \leq x < b$. The infinite interval (a, ∞) means all $x > a$, with ∞ indicating no upper bound on the right.



Composing continuous functions. Another way to combine functions $f(x)$ and $g(x)$ is to *compose* or *chain* them, taking the output of g as the input of f to obtain the new function $f(g(x))$. Composition also preserves continuity: if $g(x)$ is continuous at $x = a$, and $f(x)$ is continuous at $x = g(a)$, then $f(g(x))$ is continuous at $x = a$. This follows from the following theorem:

Composition Law: We have:

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

provided $\lim\limits_{x\to a} g(x) = b$ and $f(x)$ is continuous at $x = b$.

Proof. For any desired output error bound $\varepsilon > 0$, we must find some input accuracy $\delta > 0$ such that $|x - a| < \delta$ forces $|f(g(x)) - f(b)| < \varepsilon$.

Take any $\varepsilon > 0$. Since $f(y)$ is continuous at $y = b$, there is $\delta_1 > 0$ such that $|y - b| < \delta_1$ forces $|f(y) - f(b)| < \varepsilon$. Also, since $\lim_{x\to a} g(x) = b$, there exists $\delta > 0$ such that $0 < |x - a| < \delta$ forces $|g(x) - b| < \delta_1$. Therefore $0 < |x - a| < \delta$ forces $|g(x) - b| < \delta_1$, which in turn forces $|f(g(x)) - f(b)| < \varepsilon$, as required.

Intermediate Value Theorem:

> If $f(x)$ is continuous for x in the interval $[a, b]$, and r is between $f(a)$ and $f(b)$, meaning either $f(a) < r < f(b)$ or $f(a) > r > f(b)$, then there is a value[‡] $c \in (a, b)$ such that $f(c) = r$.

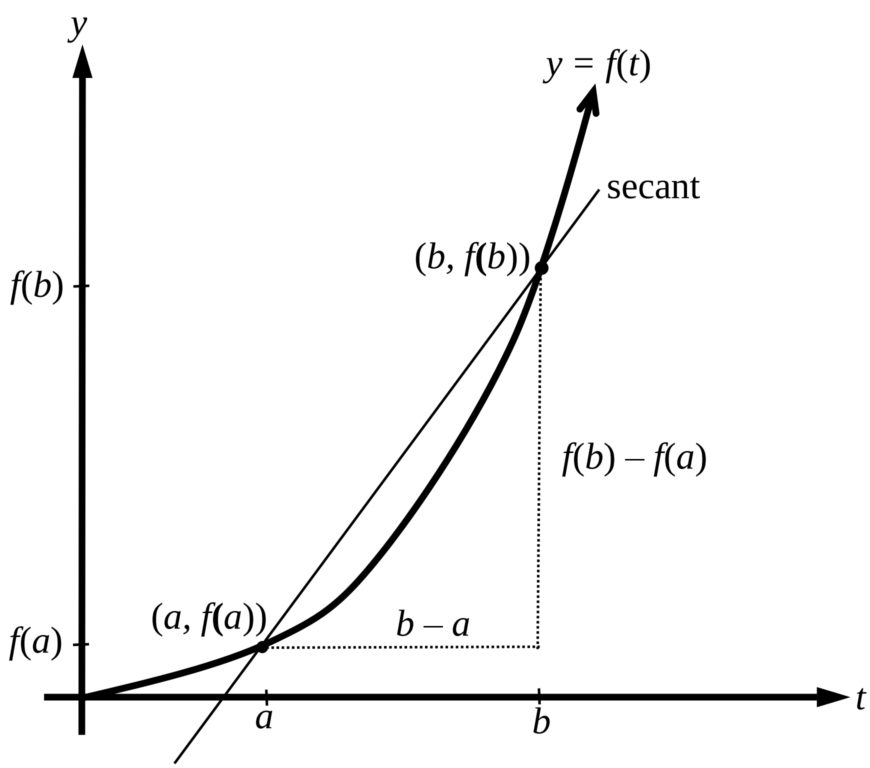
This says that as the function value $f(x)$ goes continuously from $f(a)$ to $f(b)$, perhaps rising and falling many times, it must pass through every value r between $f(a)$ and $f(b)$.

Note that this is *not* necessarily true for a discontinuous function like $g(x)$ in the graph above: taking $[a, b] = [2, 4]$, we have $g(2) \approx 1.7$, $g(4) \approx 11.2$, and $g(2) < 7 < g(4)$, but there is a vertical asymptote discontinuity at $t = 3$, and there is *no* $c \in (2, 4)$ with $g(c) = 7$.

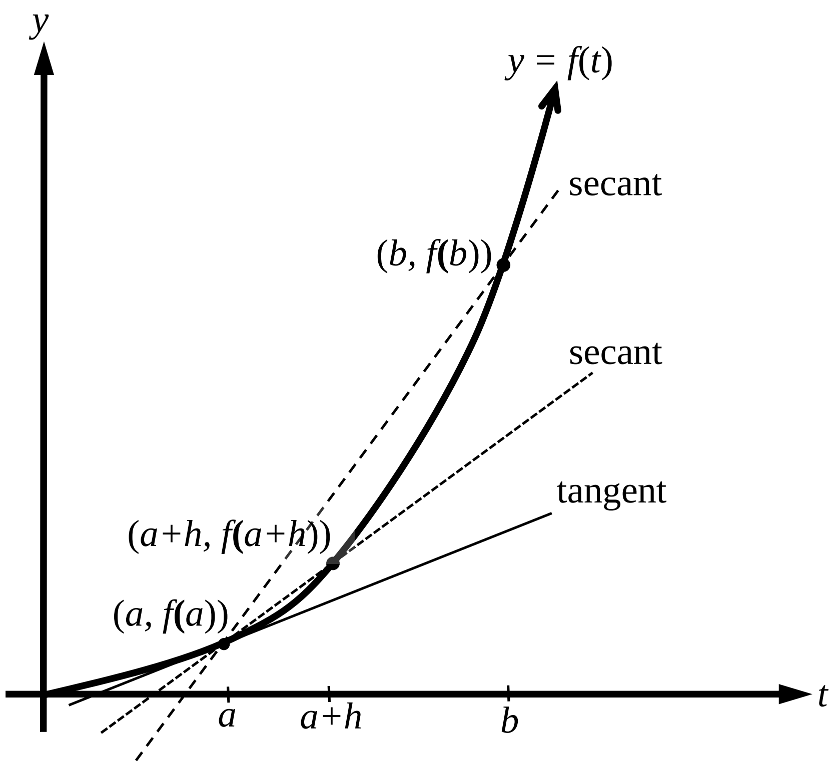
However, $g(x)$ *is* continuous over the interval $[0, 1]$, with $g(0) \approx -0.33$, $g(1) \approx 0.72$, and $g(0) < 0 < g(1)$, so the Theorem says there must be some $c \in (0, 1)$ with $g(c) = 0$. This is just the x-intercept visible in the graph.

EXAMPLE: Show that there exists a solution $x = c$ to the equation $\cos(x) = x$. We have no easy way of solving this equation, but writing $f(x) = \cos(x) - x$, we know that $f(0) = 1$, $f(\pi) = -1-\pi$, and $f(0) > 0 > f(\pi)$. Since $f(x)$ is continuous, the Theorem guarantees some $c \in (0, \pi)$ with $f(c) = 0$, meaning $\cos(c) = c$.

[‡]Here c lies in the open interval (a, b), between a and b but excluding both endpoints: $a < c < b$.



y
y = f(t)
secant
(b, f(b))
f(b)
f(b) – f(a)
(a, f(a))
b – a
f(a)
t
a
b

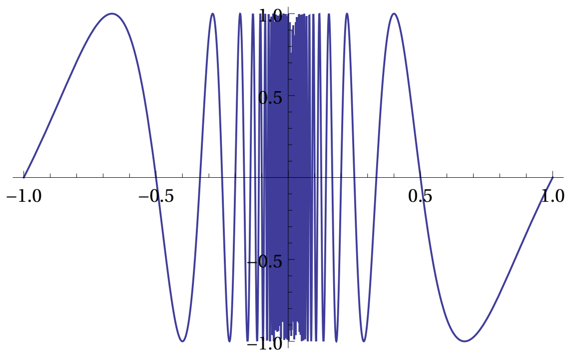

y
y = f(t)
secant
(b, f(b))
secant
tangent
(a+h, f(a+h))
(a, f(a))
t
a
a+h
b

graph sin(pi/x) for x=−1 to 1

Input interpretation:

plot	$\sin\left(\frac{\pi}{x}\right)$	$x = -1$ to 1

Plot:

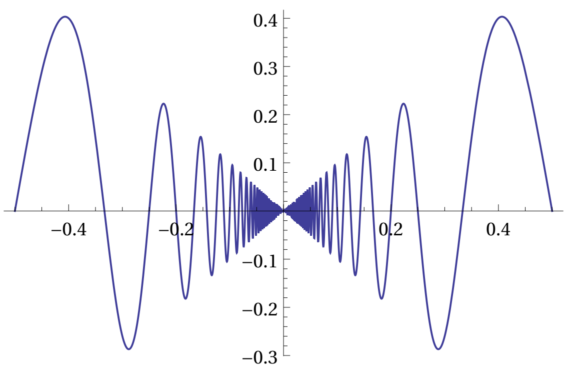


graph x sin(pi/x) for x=−.5 to .5

Input interpretation:

plot	$x \sin\left(\frac{\pi}{x}\right)$	$x = -0.5$ to 0.5

Plot:



Definition of derivative. In Notes §1.4, we saw that instantaneous velocity can be obtained as a limit of average velocities over shorter and shorter time increments. Also, the tangent slope of a graph at a given point can be obtained as a limit of secant slopes getting closer and closer to the point. Both these definitions compute a rate of change: velocity is the rate of change of position with respect to time, and slope is the rate of vertical rise with respect to horizontal run.

For any function $f(x)$, we can compute its instantaneous rate of change with respect to x in analogy with the above examples.

Definition. The *derivative* of a function $f(x)$ at $x = a$, denoted $f'(a)$, means:

$$f'(a) \;=\; \lim_{h\to 0} \frac{f(a{+}h) - f(a)}{h}\,.$$

Here $f(a{+}h) - f(a)$ is the change in $f(x)$ from $x = a$ to $x = a{+}h$, and h is the increment of x. The average rate of change over the interval* $[a, a{+}h]$ is the *difference quotient* $\frac{f(a+h)-f(a)}{h}$, and the instantaneous rate of change at $x = a$ is the limit over smaller and smaller increments, $h \to 0$.

Another way to write this is to substitute x for the endpoint of the interval, $a + h = x$, approaching a with increment $h = x - a$:

$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a}\,.$$

In graphical terms, the derivative $f'(a)$ is the slope of the tangent line which touches the graph $y = f(x)$ at the point $(a, f(a))$, and the equation of the tangent line is $y = f(a) + f'(a)(x - a)$.

When the limit $f'(a)$ exists, we say $f(x)$ is *differentiable* at $x = a$. When the limit does not exist, $f'(a)$ is undefined, and we say $f(x)$ is *non-differentiable* or *singular* at $x = a$. In this case, the function $f(x)$ does not have a well-defined rate of change at $x = a$, and the graph $y = f(x)$ does not have a single tangent line at $(a, f(a))$. (See below, *Left and right derivatives.*)

Derivatives of standard functions. A derivative is the limit of a small change in $f(x)$ divided by a small change in x, so it will always be a difficult limit of the form $\frac{0}{0}$, with no defined value if we plug in $h = 0$. To evaluate it, we must find some trick to cancel vanishing factors in the numerator and denominator, as we have seen in Notes §1.4 and §1.6 (*Limits by canceling*).

EXAMPLE: Find $f'(2)$ for $f(x) = \frac{1}{x+1}$. We compute the derivative by combining fractions over a common denominator, then canceling the vanishing factors $\frac{h}{h}$:

$$f'(2) \;=\; \lim_{h\to 0} \frac{f(2{+}h) - f(2)}{h} \;=\; \lim_{h\to 0} \frac{\frac{1}{(2+h)+1} - \frac{1}{2+1}}{h} \;=\; \lim_{h\to 0} \frac{1}{h}\left(\frac{1}{h{+}3} - \frac{1}{3}\right)$$

Notes by Peter Magyar magyar@math.msu.edu

*Note that h can also be a small negative value, in which case this is the rate of change over $[a{+}h, a]$.

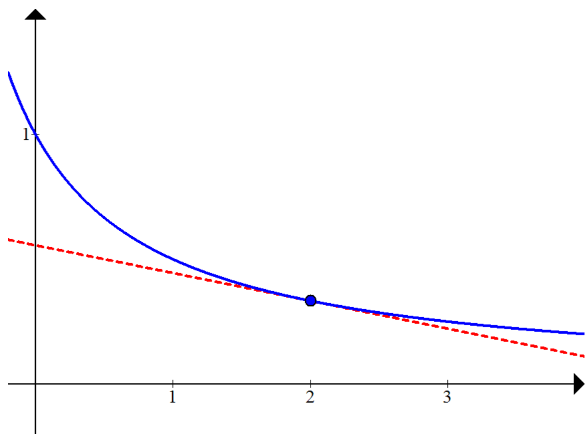
$$= \lim_{h\to 0} \frac{1}{h}\left(\frac{3-(h+3)}{3(h+3)}\right) = \lim_{h\to 0} \frac{1}{h}\left(\frac{-h}{3(h+3)}\right) = \lim_{h\to 0} \frac{-1}{3(h+3)} = \frac{-1}{3(0+3)} = -\frac{1}{9}.$$

Let us compare the same calculation with the alternative variable $x = a + h$:

$$f'(2) = \lim_{x\to 2} \frac{f(x)-f(2)}{x-2} = \lim_{x\to 2} \frac{\frac{1}{x+1}-\frac{1}{2+1}}{x-2} = \lim_{x\to 2} \frac{\frac{3-(x+1)}{3(x+1)}}{x-2}$$

$$= \lim_{x\to 2} \frac{2-x}{3(x+1)(x-2)} = \lim_{x\to 2} \frac{-1}{3(x+1)} = -\frac{1}{3(2+1)} = -\frac{1}{9},$$

where we cancel the vanishing factors $\frac{2-x}{x-2} = \frac{-(x-2)}{x-2} = -1$. Graphically, this looks like:



The curve is $y = f(x) = \frac{1}{x+1}$, and the tangent line at $(a, f(a)) = (2, \frac{1}{3})$ has slope $f'(2) = -\frac{1}{9}$, so its equation is: $y = \frac{1}{3} - \frac{1}{9}(x-2)$.

EXAMPLE: Find $f'(2)$ for $f(x) = \sqrt{x}$. Our trick is to multiply by a conjugate radical to liberate $2+h$ from under the $\sqrt{\ }$, then cancel $\frac{h}{h}$:

$$f'(2) = \lim_{h\to 0} \frac{\sqrt{2+h}-\sqrt{2}}{h} = \lim_{h\to 0} \frac{\sqrt{2+h}-\sqrt{2}}{h} \cdot \frac{\sqrt{2+h}+\sqrt{2}}{\sqrt{2+h}+\sqrt{2}}$$

$$= \lim_{h\to 0} \frac{\sqrt{2+h}^2-\sqrt{2}^2}{h(\sqrt{2+h}+\sqrt{2})} = \lim_{h\to 0} \frac{h}{h(\sqrt{2+h}+\sqrt{2})} = \frac{1}{\sqrt{2+0}+\sqrt{2}} = \frac{1}{2\sqrt{2}}.$$

Here we used the identity $(a-b)(a+b) = a^2 - b^2$ with $a = \sqrt{2+h}$ and $b = \sqrt{2}$.

Left and right derivatives. Let us find $f'(1)$ for the function defined by:

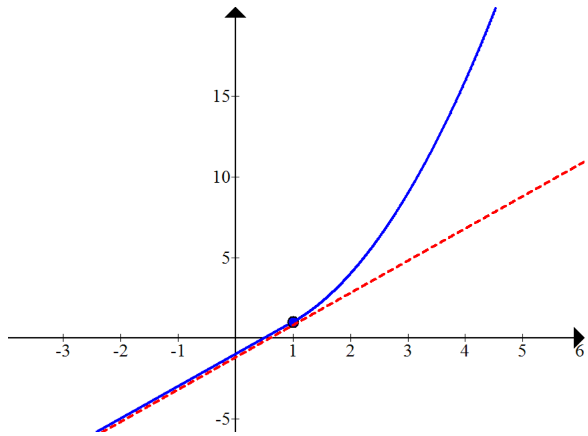
$$f(x) = \begin{cases} 2x-1 & \text{for } x \le 1 \\ x^2 & \text{for } x \ge 1. \end{cases}$$

Since the function is defined differently on the two sides of $x = 1$, we must compute one-sided derivative limits, to see if the two-sided limit exists.

$$\lim_{h\to 0^+} \frac{f(1+h)-f(1)}{h} = \lim_{h\to 0^+} \frac{(1+h)^2-1^2}{h} = \lim_{h\to 0^+} \frac{(1+2h+h^2)-1}{h} = \lim_{h\to 0^+} \frac{h(2+h)}{h} = 2.$$

$$\lim_{h\to 0^-} \frac{f(1+h)-f(1)}{h} \;=\; \lim_{h\to 0^-} \frac{2(1+h)-1-1}{h} \;=\; \lim_{h\to 0^-} \frac{2+2h-2}{h} \;=\; 2\,.$$

Since these one-sided limits agree, we have $f'(1) = \lim_{h\to 0} \frac{f(1+h)-f(1)}{h} = 2$. The graph is:

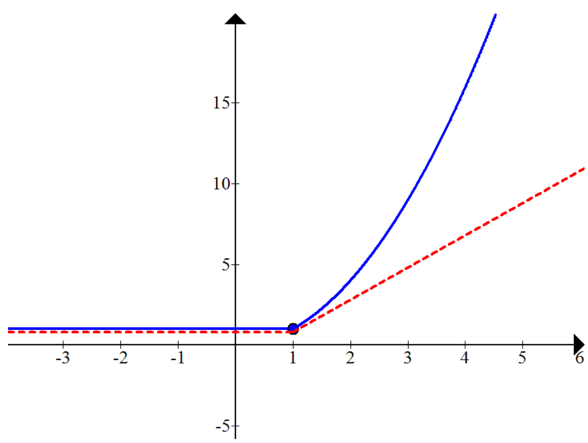


The graph clearly has a transition at $x = 1$, but it still has a well-defined tangent line.

On the other hand, if we take:

$$g(x) \;=\; \begin{cases} 1 & \text{for } x \le 1 \\ x^2 & \text{for } x \ge 1 \end{cases}$$

Then we can compute that $\lim_{h\to 0^+} \frac{g(1+h)-g(1)}{h} = 2$, but $\lim_{h\to 0^-} \frac{g(1+h)-g(1)}{h} = 0$, which means the graph has two different tangent slopes to the left and right of this point, namely it has a corner:



That is, the derivative $g'(1)$ does not exist, $g(x)$ is non-differentiable at $x = 1$, and the graph does not have a well-defined tangent line at the corner. In real-world terms, this function could model the distance fallen by an object held still, then thrown down with speed 2 at time $t = 1$. Before dropping, the speed is 0; immediately after, the speed is 2; and there is no well-defined speed at the moment of throwing. (A more detailed analysis would take into account the gradual acceleration during the throw, which would round off the corner of the graph.)

Numerical derivatives. In Notes §1.4, we computed instantaneous velocity as the derivative of the position function $f(t)$ with respect to time t. For any function which models the dependence between two real-world variables, the derivative gives the rate of change of the dependent variable with respect to the independent variable.

EXAMPLE: A rough model of atmospheric pressure P at height s is given by the function: $P = f(s) = 15c^s$, where P is in pounds per square inch (psi), s is miles above sea level, and the constant $c = 0.81$. How quickly does the pressure drop, with respect to height, at sea-level and at 2 miles up?

At sea level $s = 0$ ft, the pressure is $f(0) = 15$ psi (about half the pressure of a car tire), and the rate of change (psi of pressure change per mile of height upward) is the derivative:

$$f'(0) \;=\; \lim_{h\to 0} \frac{15c^{0+h} - 15c^0}{h} \;=\; \lim_{h\to 0} 15\,\frac{c^h - 1}{h}\,.$$

In this case, we have no algebraic trick to cancel vanishing factors, so we must be content with a numerical approximation of the difference quotient. (Since $P = f(s)$ is only an approximate model anyway, we lose nothing from this further approximation.)

h	0.1	0.01	0.001	0.0001
$15(c^h{-}1)/h$	-2.85	-3.13	-3.16	-3.16

Thus, $f'(0) \approx -3.16$ psi/mi. This is a negative rate of change because a rise in height gives a drop in pressure. For each mile upward, the pressure decreases by approximately 3.16 psi, or more accurately, a small rise like 0.1 mi would decrease pressure by about 0.316 psi.

Now at $s = 2$ mi, pressure is about $f(2) = 9.84$ psi, and we compute the rate of change:

$$f'(2) \;=\; \lim_{h\to 0} \frac{15c^{2+h} - 15c^2}{h} \;=\; \lim_{h\to 0} \frac{15c^2c^h - 15c^2}{h} \;=\; \lim_{h\to 0} 15c^2{\cdot}\frac{c^h - 1}{h} \;=\; c^2 \lim_{h\to 0} 15\,\frac{c^h - 1}{h}\,.$$

Now, $c^2 \approx 0.66$, and the second factor is the limit we approximated before. Thus, $f'(2) \approx (0.66)(-3.16) \approx -2.07$ psi/mi. That is, at an altitude of 2 mi, every 0.1 mi rise decreases the pressure by about 0.207 psi.

Derivative Function

In Notes §2.1, we defined the derivative of a function $f(x)$ at $x = a$, namely the number $f'(a)$. Since this gives an output $f'(a)$ for any input a, the derivative defines a function.

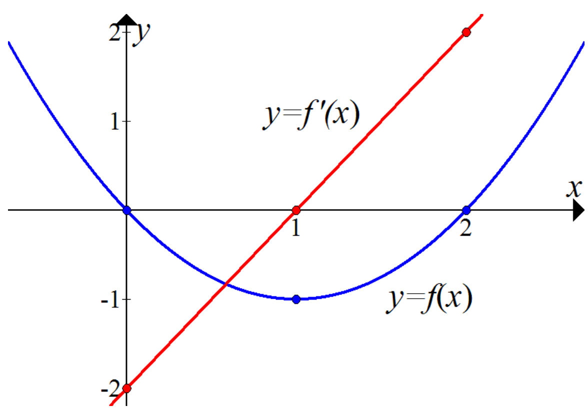
> *Definition:* For a function $f(x)$, we define the *derivative function* $f'(x)$ by:
>
> $$f'(x) \;=\; \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \;=\; \lim_{z\to x} \frac{f(z) - f(x)}{z - x}\,.$$
>
> If the limit $f'(a)$ exists for a given $x = a$, we say $f(x)$ is *differentiable* at a; otherwise $f'(a)$ is undefined, and $f(x)$ is *non-differentiable* or *singular* at a.

This just repeats the definitions in Notes §2.1, except that we think of the derivative as a function of the variable x, rather than as a numerical value at a particular point $x = a$. The choice of letters is meant to suggest different kinds of variables, but they do not have any strict logical meaning: for example, $f(x) = x^2$, $f(a) = a^2$, and $f(t) = t^2$ all define the same function, and $\lim_{x\to a} f(x) = \lim_{t\to a} f(t) = \lim_{z\to a} f(z)$ are all the same limit.

Differentiation. Another name for derivative is *differential.* When we compute $f'(x)$, we *differentiate* $f(x)$. The process of finding derivatives is *differentiation.*

As usual for mathematical objects, we can think of derivatives on four levels of meaning. The physical meaning of $f'(x)$ is the rate of change of $f(x)$ per unit change in x; for example velocity is the derivative of the position function at time t. At the end of Notes §2.1, we also saw how to compute a numerical approximation of a derivative as the diffference quotient for a small value of h (see also §2.9). In this section, we explore the geometric meaning as the slopes of the graph $y = f(x)$, and algebraic methods for computing the limit $f'(x)$.

EXAMPLE: Let $f(x) = x(x-2)$, with graph $y = f(x)$ in blue:



We can sketch the derivative graph $y = f'(x)$ in red, purely from the original graph $y = f(x)$, without any computation. The *slope* of the original graph above a given x-value is the *height* of the derivative graph above that x-value.

At the minimum $x = 1$, the original graph $y = f(x)$ is horizontal and its slope is zero, so $f'(1) = 0$, and we plot the point $(1, 0)$ on the derivative graph $y = f'(x)$. To the right of this point, $y = f(x)$ has positive slope, getting steeper and steeper; so $y = f'(x) > 0$ is above the x-axis, getting higher and higher. Above $x = 2$, the tangent of $y = f(x)$ has slope approximately 2 (considering the relative x and y scales), so we plot $(2, 2)$ on $y = f'(x)$.

Notes by Peter Magyar `magyar@math.msu.edu`

As we move left from $x = 1$, the graph $y = f(x)$ has negative slope, getting steeper and steeper, so $y = f'(x) < 0$ is below the x-axis, getting lower and lower. Above $x = 0$, we estimate $y = f(x)$ to have slope -2, and we plot $(0, -2)$ on $y = f'(x)$. Thus, $y = f'(x)$ looks like the red line in the above picture.

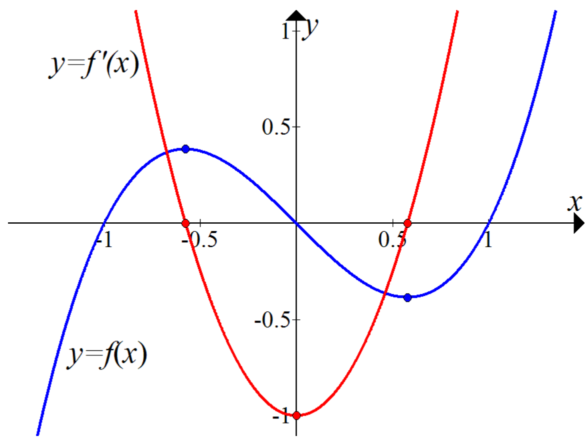
Next we differentiate algebraically. For any value of x:

$$f'(x) \;=\; \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \;=\; \lim_{h\to 0} \frac{(x+h)(x+h-2) - x(x-2)}{h}$$

$$= \lim_{h\to 0} \frac{x^2+2xh+h^2-2x-2h-x^2+2x}{h} \;=\; \lim_{h\to 0} \frac{2xh-2h+h^2}{h} \;=\; \lim_{h\to 0} 2x-2+h \;=\; 2x-2.$$

That is, $f'(x) = 2x - 2$, which agrees with our sketch of the derivative graph.

EXAMPLE: Let $f(x) = x^3 - x$, with graph in blue:



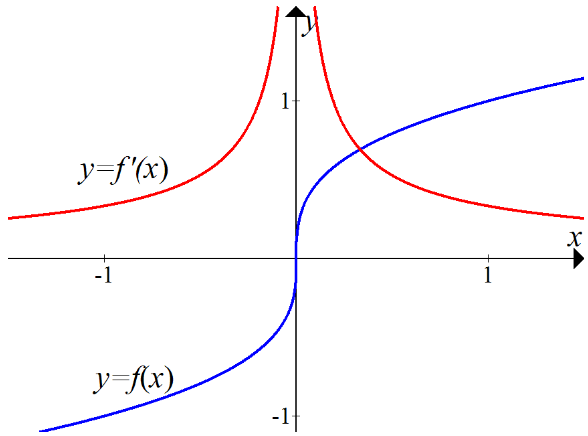
The original graph $y = f(x)$ has a valley with horizontal tangent at $x \cong 0.6$, so the derivative $f'(0.6) \cong 0$, and we plot the approximate point $(0.6, 0)$ on the derivative graph $y = f'(x)$; and similarly the hill on $y = f(x)$ corresponds to the point $(-0.6, 0)$ on $y = f'(x)$. Between these x-values, the slope of $y = f(x)$ is negative, with the slope at $x = 0$ being about -1, so $y = f'(x) < 0$ is below the x-axis, bottoming out at $(0, -1)$.

Algebraically:

$$f'(x) \;=\; \lim_{h\to 0} \frac{((x+h)^3-(x+h)) - (x^3-x)}{h} \;=\; \lim_{h\to 0} \frac{(x^3+3x^2h+3xh^2+h^3-x-h) - x^3+x}{h}$$

$$= \lim_{h\to 0} \frac{3x^2h + 3xh^2 + h^3 - h}{h} \;=\; \lim_{h\to 0} 3x^2 + 3xh + h^2 - 1 \;=\; 3x^2 - 1.$$

EXAMPLE: Let $f(x) = \sqrt[3]{x}$, the cube root function, with graph in blue:



The slopes of the original graph $y = f(x)$ are all positive, with the same slope above a given x and its reflection $-x$. Thus the derivative graph $y = f'(x) > 0$ lies above the x-axis, and it is symmetric across the y-axis (an even function). The slope of $y = f(x)$ gets smaller for large positive or negative x, and it gets steeper and steeper near the origin, with a vertical tangent at $x = 0$. Thus $y = f'(x)$ approaches the x-axis for large x, and shoots up the y-axis on both sides of $x = 0$, with $f'(0)$ undefined.

Algebraically, we have: $f'(x) = \lim_{h\to 0} \frac{\sqrt[3]{x+h}-\sqrt[3]{x}}{h}$. We must liberate $\sqrt[3]{x+h}$ from under the $\sqrt[3]{\ }$, so as to be able to cancel $\frac{h}{h}$. In Notes §2.1, we multiplied top and bottom by the conjugate radical, exploiting the identity $(a-b)(a+b) = a^2 - b^2$. Here we have cube roots, so we use the identity: $(a-b)(a^2+ab+b^2) = a^3 - b^3$, taking $a = \sqrt[3]{x+h}$ and $b = \sqrt[3]{x}$:

$$f'(x) = \lim_{h\to 0} \frac{\sqrt[3]{x+h}-\sqrt[3]{x}}{h} \cdot \frac{\sqrt[3]{x+h}^{\,2} + \sqrt[3]{x+h}\,\sqrt[3]{x} + \sqrt[3]{x}^{\,2}}{\sqrt[3]{x+h}^{\,2} + \sqrt[3]{x+h}\,\sqrt[3]{x} + \sqrt[3]{x}^{\,2}}$$

$$= \lim_{h\to 0} \frac{\sqrt[3]{x+h}^{\,3} - \sqrt[3]{x}^{\,3}}{h(\sqrt[3]{x+h}^{\,2} + \sqrt[3]{x+h}\,\sqrt[3]{x} + \sqrt[3]{x}^{\,2})} = \lim_{h\to 0} \frac{x+h-x}{h(\sqrt[3]{x+h}^{\,2} + \sqrt[3]{x+h}\,\sqrt[3]{x} + \sqrt[3]{x}^{\,2})}$$

$$= \lim_{h\to 0} \frac{1}{\sqrt[3]{x+h}^{\,2} + \sqrt[3]{x+h}\,\sqrt[3]{x} + \sqrt[3]{x}^{\,2}} = \frac{1}{\sqrt[3]{x+0}^{\,2} + \sqrt[3]{x+0}\,\sqrt[3]{x} + \sqrt[3]{x}^{\,2}} = \frac{1}{3\sqrt[3]{x}^{\,2}}\,.$$

In the Notes §2.3, we will develop standard rules for computing derivatives, which let us avoid such complicated limit calculations.

Continuity Theorem. Here is a basic fact relating derivatives and continuity:

Theorem: If $f(x)$ is differentiable at $x=a$, then $f(x)$ is also continuous at $x=a$.

Turing this around, we have the equivalent negative statement (the contrapositive): If $f(x)$ is *not* continuous at $x = a$, then it is *not* differentiable at $x = a$. That is, a discontinuity is also a non-differentiable point (a singularity).

Proof of Theorem: Assume $f(x)$ is differentiable at $x = a$, meaning $f'(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a}$ is defined. The Limit Law for Products gives:

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

Thus $0 = \lim_{h\to 0}[f(a+h) - f(a)] = [\lim_{h\to 0} f(a+h)] - f(a)$, and $\lim_{h\to 0} f(a+h) = f(a)$, showing that $f(x)$ is continuous at $x = a$.

So far, we have seen how various real-world problems (rate of change) and geometric problems (tangent lines) lead to derivatives. In this section, we will see how to solve such problems by computing derivatives (differentiating) algebraically.

Notations. We have seen the Newton notation $f'(x)$ for the derivative of $f(x)$. The alternative Leibnitz notation for the derivative is $\frac{df}{dx}$, meant to remind us of the definition of $f'(x)$ as the limit of difference quotients:

$$f'(x) \;=\; \frac{df}{dx} \;=\; \lim_{\Delta x\to 0} \frac{\Delta f}{\Delta x}\,.$$

Here $\Delta f = f(x{+}h) - f(x)$, the difference* in $f(x)$ produced by the difference $\Delta x = (x{+}h){-}x = h$. Also, df and dx are meant to suggest very small Δf and Δx, but $\frac{df}{dx}$ is not literally the quotient of two small quantities, just a complicated symbol meaning the limit of such quotients.

To illustrate: for $f(x) = x^2$, the formula $f'(x) = (x^2)' = 2x$ can be written in Leibnitz notation as:

$$\frac{df}{dx} \;=\; \frac{d}{dx}(x^2) \;=\; 2x.$$

The symbol $\frac{df}{dx}$ means the function $f'(x)$; for a particular value of a derivative at $x = a$, we write $f'(a) = \frac{df}{dx}\Big|_{x=a}$. The notation $f' = Df$ is also used, and $f'(x) = Df(x)$.

Basic Derivatives. To compute derivatives without a limit analysis each time, we use the same strategy as for limits in Notes §1.6: we establish the derivatives of some basic functions, then we show how to compute the derivatives of sums, products, and quotients of known functions.

Theorem: (i) For a constant function $f(x) = c$, we have $\frac{d}{dx}(c) = (c)' = 0$.
(ii) For $f(x) = x$, we have $\frac{d}{dx}(x) = (x)' = 1$.
(iii) For $f(x) = x^p$ with p any real number, we have:

$$\tfrac{d}{dx}(x^p) = (x^p)' = px^{p-1}.$$

Proof: (i) and (ii) follow easily from the definition of $f'(x)$. We prove (iii) in stages, for more and more general powers p, relying repeatedly on the family of formulas: $a^n - b^n = (a{-}b)(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \cdots + b^{n-1})$, valid for $n = 1, 2, 3, \ldots$ First, we consider a whole number $p = n$, and take $a = x{+}h$ and $b = x$:

$$(x^n)' \;=\; \lim_{h\to 0} \frac{(x{+}h)^n - x^n}{h} \;=\; \lim_{h\to 0} \frac{((x{+}h){-}x)\,((x{+}h)^{n-1}{+}(x{+}h)^{n-2}x{+}\cdots{+}x^{n-1})}{h}$$

$$= \lim_{h\to 0}\, (x{+}h)^{n-1}{+}(x{+}h)^{n-2}x{+}\cdots{+}x^{n-1} \;=\; (x{+}0)^{n-1}{+}(x{+}0)^{n-2}x{+}\cdots{+}x^{n-1} \;=\; nx^{n-1}.$$

Thus, $(x^n)' = nx^{n-1}$, and (iii) holds for $p = n$.

Second, we do a similar calculation for a negative integer $p = -n$, so that $x^p = \frac{1}{x^n}$; in the derivative limit, we combine fractions and apply the $a^n - b^n$ formula with $a = x$ and $b = x{+}h$. The result simplifies to $-\frac{nx^{n-1}}{x^{2n}} = (-n)x^{(-n)-1}$.

Third, we consider a fraction $p = \frac{n}{m}$ with m a whole number and n an integer, so that $x^p = x^{\frac{m}{n}} = \sqrt[m]{x^n}$. We take the derivative limit with numerator $\sqrt[m]{(x{+}h)^n} - \sqrt[m]{x^n} = a - b$.

Notes by Peter Magyar `magyar@math.msu.edu`

*Δ is capital letter delta, the Greek D, standing for "difference". The small letter delta is δ.

As in §2.2 for $\sqrt[3]{x}$, multiplying top and bottom by $a^{m-1}+a^{m-2}b+a^{m-3}b^2+\cdots+b^{m-1}$ gets rid of the radicals $\sqrt[m]{\ }$, leaving the numerator $a^m - b^m = (x{+}h)^n - x^n$, which we handled previously. Again, the limit eventually simplifies to formula (iii).

Formula (iii) is also valid for an irrational power like $p = \sqrt{2}$, but this requires more theory: we will have to wait until Calculus II to even state a clear definiton of $x^{\sqrt{2}}$.

Having computed all these limits, we never have to do so again. Just from quoting the Theorem, we get formulas like: $(x^2)' = 2x^1 = 2x$; $(x^{10})' = 10x^9$;

$$\left(\sqrt[3]{x}\right)' \;=\; (x^{1/3})' \;=\; \tfrac{1}{3}x^{-2/3} \;=\; \tfrac{1}{3\sqrt[3]{x^2}}\,; \qquad \left(\tfrac{1}{x}\right)' \;=\; (x^{-1})' \;=\; (-1)x^{-1-1} \;=\; -\tfrac{1}{x^2}\,.$$

Derivative Rules. Suppose the functions $f(x), g(x)$ are differentiable at x, so that $f'(x)$ and $g'(x)$ exist. Then we get the following derivatives:

- *Sum:* $(f(x)+g(x))' = f'(x)+g'(x)$.
- *Difference:* $(f(x)-g(x))' = f'(x)-g'(x)$.
- *Constant Multiple:* $(c\,f(x))' = c\,f'(x)$ for any constant c.
- *Product:* $(f(x)g(x))' = f'(x)g(x)+f(x)g'(x)$.
- *Quotient:* $\left(\dfrac{f(x)}{g(x)}\right)' = \dfrac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}$, where $g(x)\neq 0$.

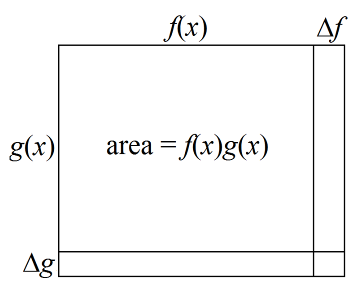
The first three of these Rules, which express the *linearity* of the derivative operation, are intuitive and easy to prove. For example the Sum Rule:

$$\begin{aligned}(f(x)+g(x))' \;&=\; \lim_{h\to 0} \tfrac{(f(x+h)+g(x+h))-(f(x)+g(x))}{h} \;=\; \lim_{h\to 0} \tfrac{f(x+h)-f(x)}{h} + \tfrac{g(x+h)-g(x)}{h} \\ &=\; \lim_{h\to 0} \tfrac{f(x+h)-f(x)}{h} + \lim_{h\to 0} \tfrac{g(x+h)-g(x)}{h} \;=\; f'(x)+g'(x).\end{aligned}$$

Here the third equality follows from the Sum Law for limits in Notes §1.6.

Warning: The derivative of a product is NOT the product of derivatives.

We obtain the correct Product Rule from a geometric model: consider a rectangle with changing sides of lengths $f(x)$ and $g(x)$ depending on some variable x, the upper left rectangle below:



The product $f(x)g(x)$ is the area, and the derivative $(f(x)g(x))'$ is the rate of change of area with respect to a change in x. Suppose small increment $\Delta x = h$ produces some positive increments $\Delta f = f(x{+}h)-f(x)$ and $\Delta g = g(x{+}h)-g(x)$ in the sides, so that the increment of area, $\Delta(f\cdot g) = f(x{+}h)g(x{+}h) - f(x)g(x)$, is the area of the three edge rectangles:†

$$\Delta(f\cdot g) \;=\; (\Delta f)\cdot g(x) + f(x)\cdot(\Delta g) + (\Delta f)\cdot(\Delta g).$$

†We can check this formula algebraically for any $f(x), f(x{+}h), g(x), g(x{+}h)$: just substitute for Δf, Δg.

To get the derivative, we divide by Δx to get the difference quotient, and send $\Delta x = h \to 0$:

$$\begin{aligned}
(f(x)g(x))' &= \lim_{\Delta x\to 0} \tfrac{\Delta(f\cdot g)}{\Delta x} = \lim_{\Delta x\to 0} \tfrac{(\Delta f)g(x)}{\Delta x} + \tfrac{f(x)(\Delta g)}{\Delta x} + \tfrac{(\Delta f)(\Delta g)}{\Delta x} \\
&= \Big(\lim_{\Delta x\to 0} \tfrac{\Delta f}{\Delta x}\Big) g(x) \;+\; f(x)\Big(\lim_{\Delta x\to 0} \tfrac{\Delta g}{\Delta x}\Big) \;+\; \Big(\lim_{\Delta x\to 0} \tfrac{\Delta f}{\Delta x}\Big)\Big(\lim_{\Delta x\to 0} \Delta g\Big) \\
&= f'(x)g(x) + f(x)g'(x) + f'(x)(0) \;=\; f'(x)g(x) + f(x)g'(x).
\end{aligned}$$

Note that the vanishing third term corresponds to the tiny bottom right rectangle.

Lastly, we prove the Quotient Rule:

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

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

Here, after putting the expression over a common denominator, we have added and subtracted the quantity $f(x)g(x)$ in the numerator, leaving the limit unchanged. Our aim is to factor the first pair and last pair of terms:

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

$$= \lim_{h\to 0} \frac{1}{g(x+h)\, g(x)} \left(\frac{f(x+h) - f(x)}{h}\, g(x) - f(x)\, \frac{g(x+h) - g(x)}{h}\right)$$

$$= \frac{1}{g(x+0)\, g(x)} \big(f'(x)g(x) - f(x)g'(x)\big) \;=\; \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

We have again used several Limit Laws from Notes §1.6. We could give another proof of the Product Rule in a very similar way.

Derivative computations. By repeatedly using these Rules, we can quickly compute the derivatives of most functions.

EXAMPLE: Find $(\sqrt{x})' = \frac{d}{dx}(\sqrt{x})$. Solution: $(\sqrt{x})' = (x^{1/2})' = \frac{1}{2}x^{(1/2)-1} = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}}$, where we used the Basic Derivative $(x^p)' = px^{p-1}$ with $p = \frac{1}{2}$.

EXAMPLE: $(\sqrt{10})' = 0$ since the derivative of *any* constant, even a complicated one, is *zero*.

EXAMPLE: For $f(x) = (5x^2+1)(\sqrt{x}-3)$, find the derivative $f'(x) = \frac{df}{dx}$:

$$\begin{aligned}
\big((5x^2{+}1)(\sqrt{x}{-}3)\big)' &= \big(5x^2{+}1\big)'\,(\sqrt{x}{-}3) + \big(5x^2{+}1\big)\,(\sqrt{x}{-}3)' && \text{by Product Rule} \\
&= \big(5(x^2)'{+}(1)'\big)\,(\sqrt{x}{-}3) + \big(5x^2{+}1\big)\,((\sqrt{x})'{-}(3)') && \text{by Sum \& Const Mult Rules} \\
&= \big(5(2x^1){+}(0)\big)\,(\sqrt{x}{-}3) + \big(5x^2{+}1\big)\,\big(\tfrac{1}{2}x^{-1/2}{-}(0)\big) && \text{by Basic Derivatives} \\
&= 10x(\sqrt{x}{-}3) + (5x^2{+}1)\tfrac{1}{2\sqrt{x}} && \text{tidying up}
\end{aligned}$$

Note how we used the derivative from the previous example, $(\sqrt{x})' = \frac{1}{2}x^{-1/2}$.

Another way to find the same derivative would be to multiply out first:

$$f(x) \;=\; (5x^2{+}1)(\sqrt{x}{-}3) \;=\; 5x^2\sqrt{x} - 15x^2 + \sqrt{x} - 3 \;=\; 5x^{5/2} - 15x^2 + x^{1/2} - 3.$$

Then we get the derivative:

$$f'(x) \;=\; 5(\tfrac{5}{2}x^{(5/2)-1}) - 15(2x^1) + \tfrac{1}{2}x^{(1/2)-1} - 0 \;=\; \tfrac{25}{2}x\sqrt{x} - 30x + \tfrac{1}{2\sqrt{x}}.$$

This agrees with our previous answer, multiplied out.

EXAMPLE: Differentiate $g(t) = \frac{t^5+1}{t\sqrt{t}}$. Solution by the Quotient Rule:

$$g'(t) \;=\; \frac{dg}{dt} \;=\; \left(\frac{t^5+1}{t\sqrt{t}}\right)' \;=\; \frac{(t^5+1)'(t\sqrt{t}) - (t^5+1)(t\sqrt{t})'}{(t\sqrt{t})^2} \;=\; \frac{(5t^4)(t\sqrt{t}) - (t^5+1)(\frac{3}{2}t^{1/2})}{t^3},$$

where we use $t\sqrt{t} = t^{3/2}$.

Solution by multiplying out: $\frac{1}{t\sqrt{t}} = t^{-3/2}$, so:

$$g(t) \;=\; (t^5+1)t^{-3/2} \;=\; t^{7/2} + t^{-3/2} \qquad \text{and} \qquad g'(t) = \tfrac{7}{2}\,t^{5/2} \;-\; \tfrac{3}{2}\,t^{1/2}.$$

EXAMPLE: A block of ice has length 10cm, width 5cm, and height 20cm. Its length and width are melting at a rate of 1cm per hour, but its height is melting at 2cm per hour (because the ground is warmer than the air). How fast is the volume decreasing?

Solution: The volume is $V = \ell wh\,\text{cm}^3$, where V, ℓ, w, h are all functions of time t. To get the rate of change, we compute the derivative using the Product Rule twice, considering $\ell wh = (\ell)(wh)$:

$$\frac{dV}{dt} \;=\; V' \;=\; (\ell wh)' \;=\; (\ell)'(wh)+(\ell)(wh)' \;=\; \ell' wh+\ell(w'h+wh') \;=\; \ell' wh+\ell w'h+\ell wh'.$$

We want the melt rate at the current time $t = 0$, and we are given: $\ell(0) = 10$ cm with $\ell'(0) = -1$ cm/hr; $w(0) = 5$ cm with $w'(0) = -1$ cm/hr; and $h(0) = 20$ cm with $h'(0) = -2$ cm/hr. Thus:

$$\begin{aligned} V'(0) \;&=\; \ell'(0)w(0)h(0) + \ell(0)w'(0)h(0) + \ell(0)w(0)h'(0) \\ &=\; (-1)(5)(20) + (10)(-1)(20) + (10)(5)(-2) \;=\; -400\,\text{cm}^3/\text{hr}. \end{aligned}$$

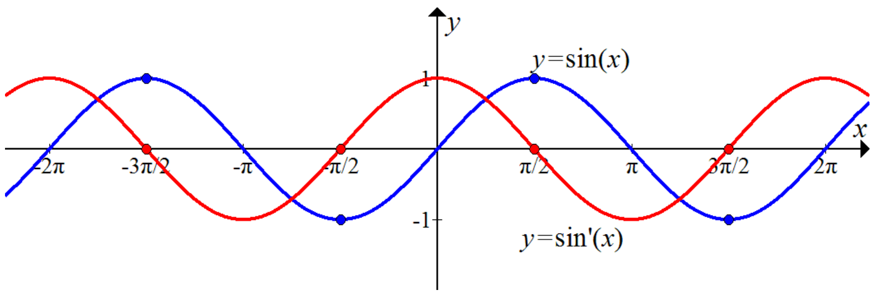
Higher derivatives. Since the derivative operation turns a function $f(x)$ into another function $f'(x)$, we can do it again to $f'(x)$, obtaining yet another function denoted $f''(x) = (f'(x))'$ or $\frac{d^2f}{dx} = \frac{d}{dx}\left(\frac{df}{dx}\right)$, called the *second derivative* of $f(x)$.

In real-world terms, if $f'(x)$ is the rate of change of $f(x)$, then $f''(x)$ is the rate of change of $f'(x)$, namely how much the rate $f'(x)$ is speeding up or slowing down.

EXAMPLE: A stone falls $f(t) = 16t^2$ ft in t seconds. Compute the repeated derivatives of this function, and interpret their physical meaning.

- The first derivative is $f'(t) = (16t^2)' = 16(2t^1) = 32t$ ft/sec. This is the *velocity* $v(t) = f'(t) = 32t$ ft/sec, increasing proportional to time.
- The second derivative is $f''(t) = (32t)' = 32$, with units ft/sec per sec = ft/sec^2. It means the rate of change of velocity, how many ft/sec of speed is gained each second. This is the *acceleration* of the stone, $a(t) = f''(t) = 32$ ft/sec^2, the constant acceleration due to gravity.
- The third derivative $f'''(t) = (32)' = 0$, meaning the rate of change of a constant acceleration is zero. The physics term for this quantity is the *jerk*, and since the jerk here is zero, we see that gravity does not jerk: it pulls smoothly. All higher derivatives are also zero; these do not have names.

Derivative of sine and cosine. The sine and cosine are important functions describing periodic motion. From the graph $y = \sin(x)$ (in blue), let us examine the slope at each point to sketch the graph of the derivative $y = \sin'(x)$ (in red), as in Notes §2.3:



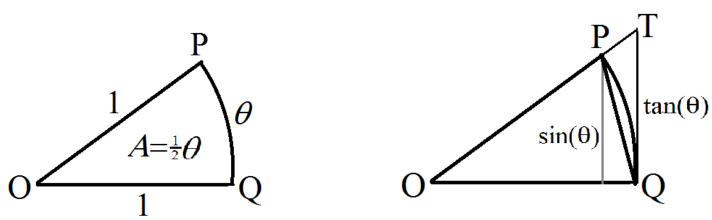
The graph $y = \sin(x)$ has hills and valleys at $x = \pm\frac{1}{2}\pi, \pm\frac{3}{2}\pi, \pm\frac{5}{2}\pi, \ldots$, so $\sin'(x) = 0$ at these points. For the interval $-\frac{1}{2}\pi < x < \frac{1}{2}\pi$, the slope of $y = \sin(x)$ is positive with a steepest slope of about 1 at $x = 0$, so $y = \sin'(x)$ swells above the x-axis from 0 to 1 to 0, and similarly on the other intervals. The graph we have drawn seems to be roughly the cosine function, so we may guess that:

$$\sin'(x) \overset{??}{=} \cos(x).$$

To prove this, we need two lemmas (minor theorems which help to prove a major one):

LEMMA: (a) $\lim_{\theta\to 0} \frac{\sin(\theta)}{\theta} = 1$ (b) $\lim_{\theta\to 0} \frac{\cos(\theta)-1}{\theta} = 0$.

Proof of (a): This is a difficult limit of the form $\frac{0}{0}$. Consider a sector OPQ of $\theta > 0$ radians: this means a pie-slice of radius $r = 1$, whose circular outer rim has length θ. (For example, $\theta = 2\pi$ would mean a full circle.)



The area of the sector is proportional to the angle, increasing from $A = 0$ for $\theta = 0$, to $A = \pi r^2 = \pi$ for $\theta = 2\pi$, so it is $A = \frac{1}{2}\theta$ for arbitrary θ. From basic trigonometry, we know that the height of the triangle $\triangle OQP$ (inside the sector) is $\sin(\theta)$, so its area is $\frac{1}{2}(\text{base})(\text{height}) = \frac{1}{2}(1)\sin(\theta) = \frac{1}{2}\sin(\theta)$. Also, the height of the triangle $\triangle OQT$ (outside the sector) is $\tan(\theta)$,* so its area is $\frac{1}{2}\tan(\theta)$. We have:

$$\begin{array}{rcccl} \text{area}(\triangle OQP) & \leq & \text{area(sector)} & \leq & \text{area}(\triangle OQT) \\ \frac{1}{2}\sin(\theta) & \leq & \frac{1}{2}\theta & \leq & \frac{1}{2}\tan(\theta) \\ \sin(\theta) & \leq & \theta & \leq & \tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}. \end{array}$$

Notes by Peter Magyar `magyar@math.msu.edu`

*The tangent function gets its name because it is the length of the segment QT tangent to the circle.

Multiplying the left inequality by $\frac{1}{\theta}$ gives: $\frac{\sin(\theta)}{\theta} \;\le\; 1$. Multiplying the right inequality by $\frac{\cos(\theta)}{\theta}$ gives: $\cos(\theta) \le \frac{\sin(\theta)}{\theta}$. Thus:

$$\cos(\theta) \;\le\; \frac{\sin(\theta)}{\theta} \;\le\; 1.$$

That is, the graph $y = \frac{\sin(\theta)}{\theta}$ is trapped between $y = \cos(\theta)$ and $y = 1$, at least for $\theta > 0$. But since $\frac{\sin(-\theta)}{(-\theta)} = \frac{\sin(\theta)}{\theta}$ and $\cos(-\theta) = \cos(\theta)$, the same inequalities hold for $\theta < 0$. But $\lim\limits_{\theta\to 0} \cos(\theta) = \cos(0) = 1$, and $\lim\limits_{\theta\to 0} 1 = 1$, so the Squeeze Theorem (§1.6) implies $\lim\limits_{\theta\to 0} \frac{\sin(\theta)}{\theta} = 1$.

Proof of (b): We compute:

$$\frac{\cos(\theta)-1}{\theta} \;=\; \frac{\cos(\theta)-1}{\theta}\cdot\frac{\cos(\theta)+1}{\cos(\theta)+1} \;=\; \frac{\cos^2(\theta)-1^2}{\theta\,(\cos(\theta)+1)}$$

$$= \frac{-\sin^2(\theta)}{\theta\,(\cos(\theta)+1)} \;=\; -\frac{\sin(\theta)}{\theta}\cdot\frac{\sin(\theta)}{\cos(\theta)+1}\,.$$

Hence by the Product Limit Law:

$$\lim_{\theta\to 0}\frac{\cos(\theta)-1}{\theta} \;=\; -\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta}\cdot\lim_{\theta\to 0}\frac{\sin(\theta)}{\cos(\theta)+1} \;=\; -1\cdot\frac{0}{1+1} \;=\; 0,$$

where we used the result established in (a).

THEOREM: $\sin'(x) = \cos(x)$ and $\cos'(x) = -\sin(x)$.

Proof: Starting with the definition of the derivative, and expanding by the Angle Addition Formula for sine, we have:

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

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

$$= \sin(x)\cdot 0 + \cos(x)\cdot 1 \;=\; \cos(x).$$

We used Lemmas (a) and (b) to get the last line. The proof of $\cos'(x) = -\sin(x)$ is similar.

General trigonometric derivatives. From these basic derivatives, we can compute the derivative of any trig function or combination of trig functions.

EXAMPLE: Compute the derivative of $\tan(x)$. The Quotient Rule for derivatives (Notes §2.3) gives:

$$\tan'(x) \;=\; \left(\frac{\sin(x)}{\cos(x)}\right)' \;=\; \frac{\sin'(x)\cos(x)-\sin(x)\cos'(x)}{\cos^2(x)}$$

$$= \frac{\cos(x)\cos(x)-\sin(x)(-\sin(x))}{\cos^2(x)} \;=\; \frac{1}{\cos^2(x)} \;=\; \sec^2(x),$$

since $\cos^2(x)+\sin^2(x) = 1$. In fact, we get the following derivatives:

$f(x)$	$\sin(x)$	$\cos(x)$	$\tan(x)$	$\sec(x)$	$\csc(x)$	$\cot(x)$
$f'(x)$	$\cos(x)$	$-\sin(x)$	$\sec^2(x)$	$\tan(x)\sec(x)$	$-\cot(x)\csc(x)$	$-\csc^2(x)$

Warning: These formulas are for angle x in **radians**, NOT in degrees (see §2.5 end).

Limits of quotients. We can also compute trigonometric limits of the form $\frac{0}{0}$. The trick is to manipulate the numerators and denominators to get factors of the form $\frac{\sin(g(x))}{g(x)}$, where $g(x)$ is any quantity which goes to zero.

EXAMPLE: Compute $\lim_{x\to 0} \frac{\sin(3x)}{x}$. We have:

$$\lim_{x\to 0} \frac{\sin(3x)}{x} \;=\; \lim_{x\to 0} \frac{\sin(3x)}{3x} \cdot \frac{3x}{x} \;=\; \lim_{x\to 0} \frac{\sin(3x)}{3x} \cdot \lim_{x\to 0} \frac{3x}{x} \;=\; \lim_{h\to 0} \frac{\sin(h)}{h} \cdot \lim_{x\to 0} 3 \;=\; 1\cdot 3 \;=\; 3.$$

Here we use $\lim_{x\to 0} \frac{\sin(3x)}{3x} = \lim_{h\to 0} \frac{\sin(h)}{h} = 1$, where we substitute[†] $h = g(x) = 3x$, so that $x \to 0$ forces $h \to 0$.

EXAMPLE: Compute $\lim_{x\to 0} \frac{\tan(x)}{\sin(\sqrt{x})}$. Starting with $\tan(x) = \frac{\sin(x)}{\cos(x)}$, we get:

$$\begin{aligned}
\lim_{x\to 0} \frac{\tan(x)}{\sin(\sqrt{x})} \;&=\; \lim_{x\to 0} \frac{1}{\cos(x)} \cdot \sin(x) \cdot \frac{1}{\sin(\sqrt{x})} \\
&=\; \lim_{x\to 0} \frac{1}{\cos(x)} \cdot \frac{\sin(x)}{x} \cdot x \cdot \frac{\sqrt{x}}{\sin(\sqrt{x})} \cdot \frac{1}{\sqrt{x}} \\
&=\; \lim_{x\to 0} \frac{\sqrt{x}}{\cos(x)} \cdot \frac{\sin(x)}{x} \cdot \frac{1}{\frac{\sin(\sqrt{x})}{\sqrt{x}}} \;=\; \frac{\sqrt{0}}{\cos(0)} \cdot 1 \cdot \frac{1}{1} \;=\; 0\,,
\end{aligned}$$

where $\lim_{x\to 0} \frac{\sin(\sqrt{x})}{\sqrt{x}} = 1$ by the substitution $h = g(x) = \sqrt{x}$.

[†]By the Limit Substitution Theorem at the end of Notes §1.7.

Chain of functions. On a Ferris wheel, your height H (in feet) depends on the angle θ of the wheel (in radians): $H = 100 + 100\sin(\theta)$. The wheel is turning at one revolution per minute, meaning the angle at t minutes is $\theta = 2\pi t$ radians. At $t = \frac{1}{12}$, we have $\theta = \frac{\pi}{6}$ and:

$$H \;=\; 100 + 100\sin(2\pi t) \;=\; 100 + 100\sin(\tfrac{\pi}{6}) = 150 \text{ ft.}$$

At this moment, how fast are you rising (in ft/min)?

The answer is given by the *Chain Rule*, which computes the derivative for a chain of functional dependencies: one variable H depends on a second variable θ, which depends on a third variable t. The Rule states:

$$\frac{dH}{dt} \;=\; \frac{dH}{d\theta}\cdot\frac{d\theta}{dt}$$

$$\frac{\text{ft}}{\text{min}} \;=\; \frac{\text{ft}}{\text{rad}}\cdot\frac{\text{rad}}{\text{min}}$$

The rate of change of height with respect to angle is:

$$\frac{dH}{d\theta} \;=\; \frac{d}{d\theta}(100 + 100\sin(\theta)) \;=\; 0 + 100\sin'(\theta)$$

$$=\; 100\cos(\theta) \;=\; 100\cos(\tfrac{\pi}{6}) \;\cong\; 86.6\,\frac{\text{ft}}{\text{rad}}.$$

The rate of change of angle with respect to time is:

$$\frac{d\theta}{dt} \;=\; \frac{d}{dt}(2\pi t) \;=\; 2\pi \;\cong\; 6.28\,\frac{\text{rad}}{\text{min}}.$$

Thus, the Chain Rule says the rate of change of height with respect to time is the product:

$$\frac{dH}{dt} \;\cong\; 86.6\frac{\text{ft}}{\text{rad}} \times 6.28\frac{\text{rad}}{\text{min}} \;\cong\; 544\frac{\text{ft}}{\text{min}}.$$

Your rate of rise is about 544 feet per minute, at time $t = \frac{1}{12}$.

Chain Rule: Let y, u, x be variables related by $y = f(u)$ and $u = g(x)$, so that $y = f(g(x))$. Then, in Leibniz notation:

$$\frac{dy}{dx} \;=\; \frac{dy}{du}\cdot\frac{du}{dx}$$

or in Newton notation:

$$f(g(x))' \;=\; f'(g(x))\cdot g'(x).$$

This holds at any value of x where $g'(x)$ and $f'(g(x))$ are both defined.

Notes by Peter Magyar `magyar@math.msu.edu`

The function $f(g(x))$ is called the *composition* of f following g, sometimes denoted $f \circ g$, so that we may write $f(g(x))'$ as $(f \circ g)'(x)$.

Proof.[*] First we assume that the value $g(a)$ is different from all other nearby output values $g(x)$: that is, for x close enough (but unequal) to a, we have $g(x) \neq g(a)$. Then we compute, using the alternative definition of derivative:

$$\begin{aligned}
(f \circ g)'(a) \;&=\; \tfrac{d}{dx} f(g(x))\Big|_{x=a} \;=\; \lim_{x \to a} \frac{f(g(x)) - f(g(a))}{x - a} \\
&= \lim_{x \to a} \frac{f(g(x)) - f(g(a))}{g(x) - g(a)} \cdot \lim_{u \to a} \frac{g(x) - g(a)}{x - a} \\
&= \lim_{u \to g(a)} \frac{f(u) - f(g(a))}{u - g(a)} \cdot \lim_{u \to a} \frac{g(x) - g(a)}{x - a} \\
&= f'(g(a)) \cdot g'(a)\,.
\end{aligned}$$

Here we used the Limit Substitution Theorem from Notes §1.7, substituting u for $g(x)$ so that $x \to a$ forces $u \to g(a)$. (Since $g(x)$ is differentiable at $x = a$, it is also continuous.)

Finally, if there is a sequence of inputs $x_1, x_2, \dots \to a$ with $g(x_i) = g(a)$, then we clearly have $g'(a) = 0$, and the right side of our formula becomes $f'(g(a)) \cdot g'(a) = 0$. On the left side, we have values $(f(g(x_i)) - f(g(a)))/(x_i - a) = 0$, which is consistent with the desired limit $(f \circ g)'(a) = 0$, and the previous argument is still valid when restricted to the set of x with $g(x) \neq g(a)$.

Differentiation Rules. Along with our previous Derivative Rules from Notes §2.3, and the Basic Derivatives from Notes §2.3 and §2.4, the Chain Rule is the last fact needed to compute the derivative of any function defined by a formula.

EXAMPLE: Find the derivative of $(x+\frac{1}{x})^{10}$. First, we use Leibnitz notation: let $y = u^{10}$ and $u = x + \frac{1}{x}$, so that $y = (x+\frac{1}{x})^{10}$. Then:

$$\begin{aligned}
\frac{dy}{dx} \;&=\; \frac{dy}{du} \cdot \frac{du}{dx} \;=\; \frac{d}{du}(u^{10}) \cdot \frac{d}{dx}\left(x+\tfrac{1}{x}\right) \;=\; 10u^9 \cdot \frac{d}{dx}\left(x+x^{-1}\right) \\
&= 10\left(x+\tfrac{1}{x}\right)^9 \cdot \left(1+(-1x^{-2})\right) \;=\; 10\left(x+\tfrac{1}{x}\right)^9\left(1-\tfrac{1}{x^2}\right).
\end{aligned}$$

Next, we redo this in Newton notation, without introducing new letters y, u. Let $f(x) = x^{10}$ with $f'(x) = 10x^9$, and $g(x) = x + \frac{1}{x} = x + x^{-1}$ with $g'(x) = 1 - x^{-2} = 1 - \frac{1}{x^2}$, so that:

$$f(g(x))' \;=\; f'(g(x)) \cdot g'(x) \;=\; 10\left(x+\tfrac{1}{x}\right)^9\left(1-\tfrac{1}{x^2}\right).$$

A third way (the quickest in practice) is to think of the composite function as an outside function $out = (\ \ )^{10}$ wrapped around an inside function $in = x + \frac{1}{x}$, so the Chain Rule becomes:

$$out(in)' \;=\; out'(in) \cdot in'$$

[*]For another proof based on linear approximations, see the Stewart text §2.5, p. 153.

Here $out' = 10(\ \ )^9$, so:

$$out(in)' \;=\; 10\big(x+\tfrac{1}{x}\big)^9 \cdot \big(x+\tfrac{1}{x}\big)' \;=\; 10\big(x+\tfrac{1}{x}\big)^9 \cdot \big(1-\tfrac{1}{x^2}\big)$$

EXAMPLE: For any function $u = g(x)$, and any number n, we have:

$$\tfrac{d}{dx}(u^n) = nu^{n-1}\,\tfrac{du}{dx} \qquad \text{and} \qquad (g(x)^n)' = n\,g(x)^{n-1}g'(x).$$

EXAMPLE: Find the derivative of $\frac{1}{\sqrt{x\cos(x)}}$. Here the outer function is $out = \frac{1}{\sqrt{\ }} = (\ \ )^{-1/2}$ with $out' = -\frac{1}{2}(\ \ )^{-3/2}$. Thus:

$$\left(\frac{1}{\sqrt{x\cos(x)}}\right)' \;=\; -\tfrac{1}{2}(x\cos(x))^{-3/2}\cdot(x\cos(x))'$$

$$=\; -\tfrac{1}{2}(x\cos(x))^{-3/2}\cdot((x)'\cos(x)+x\cos'(x)) \;=\; -\tfrac{1}{2}(x\cos(x))^{-3/2}(\cos(x)-x\sin(x))$$

Here we used the Chain Rule, then the Product Rule.

EXAMPLE: Compare the derivatives of $\sin(x^2)$ and $\sin^2(x)$. Note that if $f(x) = \sin(x)$ and $g(x) = x^2$, we have $\sin(x^2) = f(g(x))$, but $\sin^2(x) = g(f(x))$. Thus:

$$\big(\sin(x^2)\big)' \;=\; \sin'(x^2)\cdot(x^2)' \;=\; \cos(x^2)\cdot 2x \;=\; 2x\cos(x^2)$$

$$\big(\sin^2(x)\big)' \;=\; \big((\sin(x))^2\big)' \;=\; 2(\sin(x))\cdot\sin'(x) \;=\; 2\sin(x)\cos(x)$$

EXAMPLE: Find the derivative of $\sin\!\Big(\tan\!\Big(\frac{x}{x+1}\Big)\Big)$, a composition of three functions. We start by applying the Chain Rule to the outermost function $\sin(\ \ )$, with inner function $\tan\!\Big(\frac{x}{x+1}\Big)$; then we use the Chain Rule again on this.

$$\Big(\sin\Big(\tan\Big(\tfrac{x}{x+1}\Big)\Big)\Big)' \;=\; \sin'\Big(\tan\Big(\tfrac{x}{x+1}\Big)\Big)\cdot\Big(\tan\Big(\tfrac{x}{x+1}\Big)\Big)'$$

$$=\; \sin'\Big(\tan\Big(\tfrac{x}{x+1}\Big)\Big)\cdot\tan'\Big(\tfrac{x}{x+1}\Big)\cdot\Big(\tfrac{x}{x+1}\Big)'$$

$$=\; \sin'\Big(\tan\Big(\tfrac{x}{x+1}\Big)\Big)\cdot\tan'\Big(\tfrac{x}{x+1}\Big)\cdot\tfrac{(x)'(x+1)-x(x+1)'}{(x+1)^2}$$

$$=\; \cos\Big(\tan\Big(\tfrac{x}{x+1}\Big)\Big)\cdot\sec^2\Big(\tfrac{x}{x+1}\Big)\cdot\tfrac{(x+1)-x}{(x+1)^2}$$

The last factor uses the Quotient Rule.

EXAMPLE: What if we apply the Chain Rule to a complicated constant like π^3, where we consider x^3 as the outside function and the constant function $p(x) = \pi$ as the inside? Then:

$$\big(\pi^3\big)' \;=\; 3\pi^2\cdot(\pi)' \;=\; 3\pi^2\cdot 0 \;=\; 0,$$

since $(\pi)' = c' = 0$. *Any* expression with no variable in it is constant, with derivative *zero*.

Degrees versus radians. In higher mathematics, we always use radian measure (full circle $= 2\pi$ radians)†, so that $\sin(x)$ always means sine of x radians. This is essential to get the formula $\sin'(x) = \cos(x)$.

The sine with input x in degrees (full circle = 360 deg) is actually a different function, which we can denote as $\sin_{deg}(x)$. Remember that a function is a rule which converts input numbers to output numbers: it does not know that we interpret some numbers as angles, or what their units should be. Since $\sin(x)$ and $\sin_{deg}(x)$ produce different outputs from a given number x, they are different functions. In fact, we have:

$$\sin_{deg}(x) = \sin\big(\tfrac{2\pi}{360}x\big)\,.$$

The inside operation converts x from degrees to radians, then feeds this into the ordinary (radian) sine function.

This makes a crucial difference in the derivative:

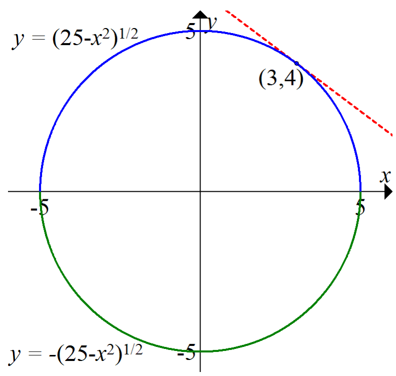
$$\sin'_{deg}(x) \;=\; \Big(\sin\big(\tfrac{2\pi}{360}x\big)\Big)' \;=\; \cos\big(\tfrac{2\pi}{360}x\big)\cdot\big(\tfrac{2\pi}{360}x\big)' \;=\; \cos_{deg}(x)\cdot\big(\tfrac{2\pi}{360}\big)\,.$$

This is why we stay away from degree measure in calculus!

†The geometric definition of radian measure is that an arc of length x on a unit circle makes an angle of x radians. The full circle, whose arc length is the circumference 2π, measures as 2π radians.

Explicit versus implicit functions. Given the circle defined by the equation $x^2+y^2=25$, suppose we wish to find the tangent line at the point $(x,y)=(3,4)$. Calculus finds a tangent slope of a function graph $y=f(x)$ as a derivative $f'(a)=\frac{df}{dx}|_{x=a}$; but there is no function specified in our problem.

Rather, we must interpret x as an independent variable, which *implicitly* makes y a function of x: to make this *explicit*, we solve the equation for y, giving $y=\pm\sqrt{25-x^2}$. That is, the circle is the union of two function graphs, $y=\sqrt{25-x^2}$ and $y=-\sqrt{25-x^2}$, each over the domain $x\in[-5,5]$.



The given point $(3,4)$ is on the top graph, and we differentiate its explicit function:

$$\frac{dy}{dx} \;=\; \frac{d}{dx}\sqrt{25-x^2} \;=\; \frac{d}{dx}(25-x^2)^{\frac{1}{2}} \;=\; \tfrac{1}{2}(25-x^2)^{-\frac{1}{2}}\,\frac{d}{dx}(25-x^2) \;=\; \frac{-x}{\sqrt{25-x^2}}.$$

Here we used the Chain Rule with outside function $(\quad)^{1/2}$. At our point, we have tangent slope $\frac{dy}{dx}|_{x=3}=y'(3)=\frac{-3}{\sqrt{25-3^2}}=-\frac{3}{4}$, and the tangent line $y=-\frac{3}{4}(x-3)+4$.

Implicit differentiation is a smoother way to do this problem. Instead of solving the equation for y, we assume $y=y(x)$ for some unkown function $y(x)$ which satisfies the equation $x^2+y(x)^2=25$. Then we differentiate both sides using the Rules:

$$\begin{aligned}\big(x^2+y(x)^2\big)' &= (25)' \\ \big(x^2\big)'+\big(y(x)^2\big)' &= 0 \\ 2x+2y(x)y'(x) &= 0.\end{aligned}$$

Note that $(x^2)'=2x$ is a Basic Derivative, but for $(y^2)'$, we need the Chain Rule with outside function $(\quad)^2$ and inside function $y=y(x)$. The derivative $y'(x)$ is the unknown we are trying to find, and now we can solve for it: $y'(x)=-\frac{x}{y(x)}$, which was easier than solving for the original $y(x)$. Since are considering the point $(x,y)=(3,4)$, we must have $y(3)=4$, so that $\frac{dy}{dx}|_{x=3}=y'(3)=-\frac{3}{y(3)}=-\frac{3}{4}$, as before.

Note that the formula $y'(x)=-\frac{x}{y(x)}$, or in Leibnitz notation $\frac{dy}{dx}=-\frac{x}{y}$, is valid for both of the functions defining the upper and lower half-circles. Since both functions obey the original equation, they both obey the derivative equation. For example, at $(x,y)=(3,-4)$, the slope is $y'(3)=-\frac{3}{y(3)}=-\frac{3}{-4}=\frac{3}{4}$.

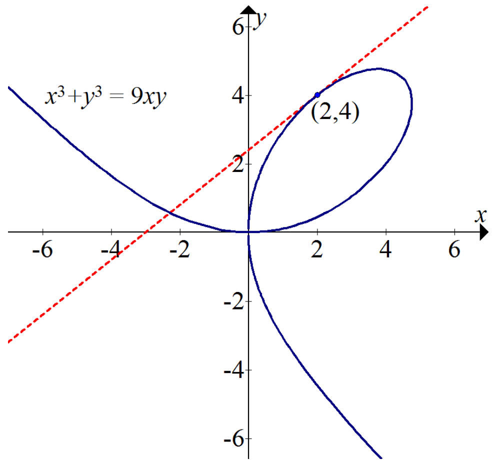
We could even take this one step further to find the second derivative implicitly:

$$y''(x) \;=\; (y'(x))' \;=\; (-\tfrac{x}{y})' \;=\; -\frac{(x)'y-x\,y'}{y^2} \;=\; -\frac{y-x(-\frac{x}{y})}{y^2} \;=\; -\frac{y^2+x^2}{y^3} \;=\; -\frac{25}{y^3}\,.$$

We used the Quotient Rule, the previous $y'=-\frac{x}{y}$, and the original equation $x^2+y^2=25$.

Notes by Peter Magyar `magyar@math.msu.edu`

Folium of Descartes. This is a curious curve discovered by the famous mathematician who gave us Cartesian xy-coordinates. It is defined by the equation: $x^3 + y^3 = 9xy$:*



We want to find the tangent line at the point $(x, y) = (2, 4)$, which is on the curve because $2^3 + 4^3 = 9(2)(4)$. In this case, there is no easy way to solve for y to get an explicit function $y(x)$; indeed, over $x \in [0, \frac{9}{2}]$, the curve is the union of *three* function graphs.

Nevertheless, implicit differentiation works without a hitch: we assume $y = y(x)$ is some unknown function which satisfies the equation, and differentiate both sides (this time in Leibnitz notation):

$$\begin{aligned}
\tfrac{d}{dx}\big(x^3 + y^3\big) &= \tfrac{d}{dx}(9xy) \\
\tfrac{d}{dx}(x^3) + \tfrac{d}{dx}(y^3) &= 9\big(\tfrac{d}{dx}(x)\, y \;+\; x\, \tfrac{d}{dx}(y)\big) \\
3x^2 + 3y^2 \tfrac{dy}{dx} &= 9y + 9x \tfrac{dy}{dx}.
\end{aligned}$$

Here we used the Sum and Product Rules, then the Chain Rule. Solving for $\frac{dy}{dx}$:

$$3y^2 \tfrac{dy}{dx} - 9x \tfrac{dy}{dx} \;=\; 9y - 3x^2, \qquad \tfrac{dy}{dx} \;=\; \frac{9y - 3x^2}{3y^2 - 9x}.$$

We do not know $y(x)$ explicitly, but our given point $(x, y) = (2, 4)$ means that $y(2) = 4$, so:

$$\left.\tfrac{dy}{dx}\right|_{x=2} \;=\; \frac{9y - 3x^2}{3y^2 - 9x} \;=\; \frac{9(4) - 3(2^2)}{3(4^2) - 9(2)} \;=\; \tfrac{4}{5}.$$

Thus, the tangent line through the point $(2, 4)$ is: $y = \frac{4}{5}(x - 2) + 4$.

Method for implicit differentiation. Given an equation involving variables x and y, we assume x is an independent variable and $y = y(x)$ is a dependent variable. To find the derivative $\frac{dy}{dx}$:

1. Take the derivative of both sides of the equation, using the Chain Rule for expressions involving $y = y(x)$ as the inside function.

2. Solve the derivative equation for the unknown $\frac{dy}{dx}$, in terms of x and y.

3. To get a specific value $y'(a) = \frac{dy}{dx}|_{x=a}$, plug in the known values $x = a$ and $y = y(a)$.

---

*To find points satisfying this equation, substitute $y = tx$ for a new variable t, and solve for x, giving: $x = \frac{9t}{1+t^3}$ and $y = \frac{9t^2}{1+t^3}$. Then each value of t gives a point (x, y) on the curve: this is called a *parametrization*.

Conceptual levels. Mathematics solves problems partly with technical tools like the differentiation rules, but its most powerful method is to translate between different levels of meaning, transforming the problems to make them accessible to our tools. Problems often originate at the physical or geometric levels, and we translate to the numerical or algebraic levels to solve them, then we translate the answer back to the original level.

Our key concept so far has been the *derivative*, with the following meanings:

- Physical: For a function $y = f(x)$, the derivative $\frac{dy}{dx} = f'(x)$ is the rate of change of y with respect to x, near a particular value of x. For a a particular input, $f'(a)$ means how fast $f(x)$ changes from $f(a)$ per unit change in x away from a. This is the main importance of derivatives.

- Geometric: For a graph $y = f(x)$, the derivative $f'(a)$ is the slope of the tangent line at the point $(a, f(a))$.

- Numerical: We approximate the derivative by the difference quotient:

$$f'(a) \cong \frac{\Delta f}{\Delta x} = \frac{f(a+h)-f(a)}{h}.$$

  The right side is the average rate of change of $f(x)$ from $x = a$ to $x = a+h$, for some small increment such as $h = 0.1$. As $\Delta x = h \to 0$, the difference quotient approaches the instantaneous rate of change, the derivative $f'(a)$.

- Algebraic: We can easily compute the derivative of almost any function defined by a formula. Basic Derivatives like $(x^p)' = px^{p-1}$, $\sin'(x) = \cos(x)$, and $\cos'(x) = -\sin(x)$ are combined using the Sum, Product, Quotient, and Chain Rules for Derivatives. Occasionally, we must go back to the definition $f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$.

Functions of motion. We consider the basic physical quantities describing motion. These are all functions of time t. (See end of §2.3.)

- *Position* or *displacement* s, the distance of an object past a reference point, in feet, at time t seconds.

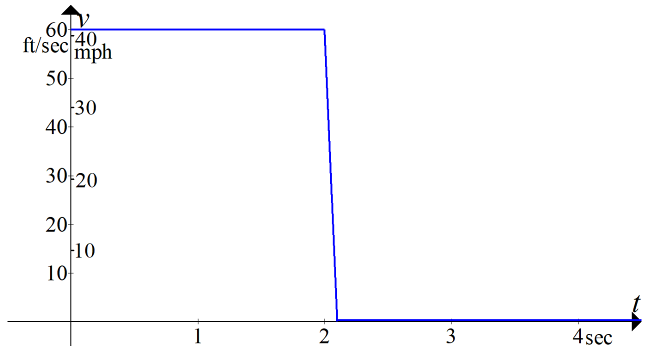
- *Velocity* $v = \frac{ds}{dt}$ or $v(t) = s'(t)$, how fast the position is increasing per second (ft/sec); this is negative if position is decreasing. The *speed* is the magnitude $|v|$.

- *Acceleration* $a = \frac{dv}{dt} = \frac{d^2s}{dt^2}$ or $a(t) = v'(t) = s''(t)$, how fast the velocity is increasing, the number of ft/sec gained each second (ft/sec^2). Equivalently, this is how fast the object is speeding up (positive) or slowing down (negative). A practical unit is the *gee* ≈ 32 ft/sec^2, the acceleration due to gravity in freefall (near the Earth's surface).

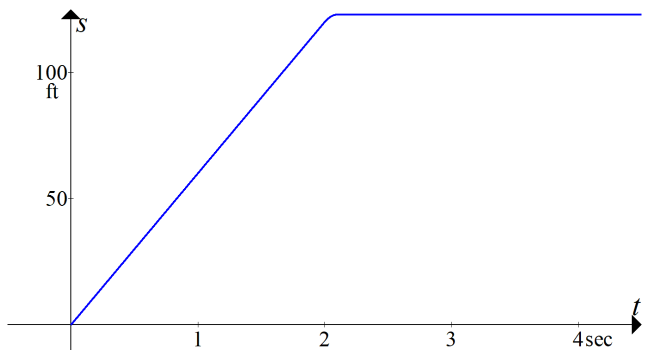
- *Jerk* $j = \frac{da}{dt} = \frac{d^3s}{dt^3}$ or $j(t) = a'(t) = s'''(t)$, rate of change of acceleration (ft/sec^3).

Notes by Peter Magyar `magyar@math.msu.edu`

Car stopping. Consider the following velocity data from a car's spedometer.

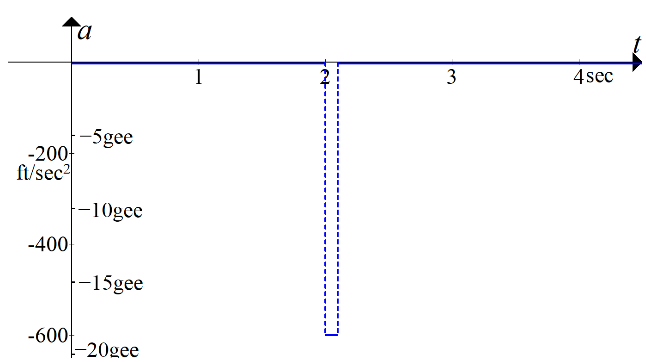


The velocity $v(t) = s'(t)$ is the derivative of the distance driven or odometer reading $s(t)$, so the *level* of the velocity graph is the *slope* of the distance graph. Thus $s(t)$ has constant slope 60 ft/sec until $t = 2$, then turns suddenly to zero slope, meaning the car has stopped.



The acceleration is the derivative $a(t) = v'(t)$, so the *slope* of the velocity graph is the *level* of the acceleration graph: this slope is zero, except at the very steep transition between 2 and 2.1 sec, which makes an average slope of:

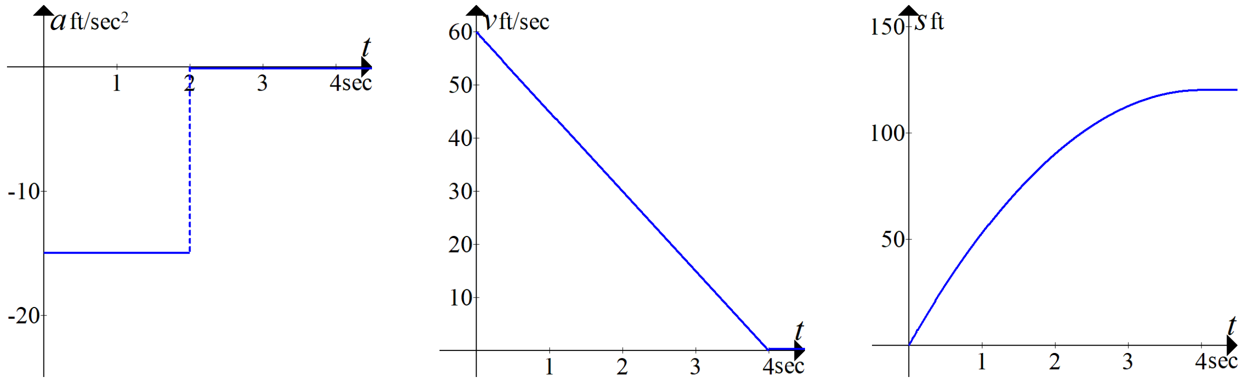
$$a \;\approx\; \frac{\Delta v}{\Delta t} \;=\; \frac{0-60}{2.1-2} \;=\; -600\text{ft/sec}^2.$$



What physical story do these graphs tell? The car went from about 40 mph to zero in 0.1 sec. Brakes cannot decelerate so quickly: this is a hard crash. The deceleration graph is zero for most of the time, but for a split second it is almost 20 gees, twenty times your weight pressing you into the seatbelt.

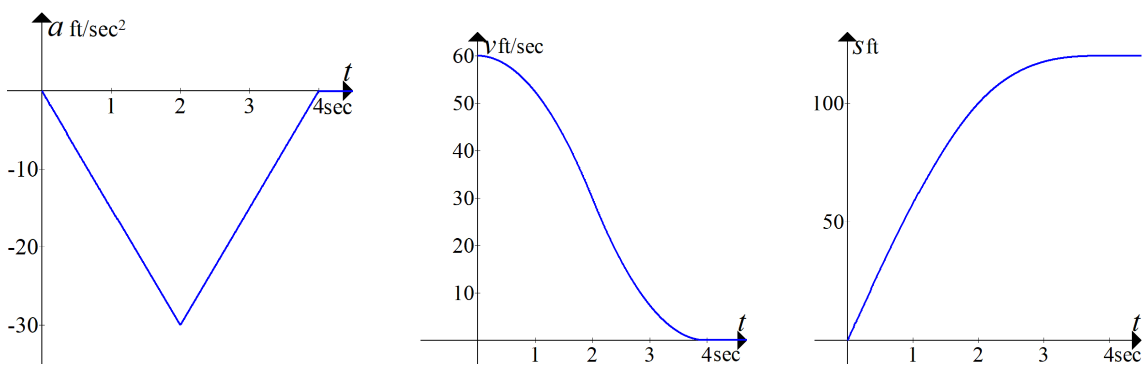
In this analysis, we translated conceptually from graphical (geometric) to physical; and also (for the gee calculation) from graphical to numerical to physical.

Braking techniques. Now imagine braking steadily at a traffic light, slowing at a constant rate until you reach a full stop. This time the deceleration is: $\frac{\Delta v}{\Delta t} = \frac{0-60}{4-0} = -15\,\text{ft/sec}^2$, less than half a gee, but still a pretty hard stop:



Even though the deceleration is not very great, it changes suddenly (instantaneously in our picture), so the derivative of acceleration (the jerk) is very large for a split second, giving a noticeable jerk or jolt at the moment of stopping, not dangerous but annoying.

Is there a braking technique which will eliminate the jerk? To prevent the sudden change in acceleration, squeeze the brake slowly down, then let it slowly up:



Now the stopping time of 4 sec requires a peak deceleration of $-30\ \text{ft/sec}^2 \approx 1$ gee, which would send the car skidding helter-skelter. You would need double the time to do this technique safely, starting to brake much earlier.

Ballistic equation. This is the formula giving the height $s(t)$ for a projectile (cannon ball) launched from initial height s_0, straight upward with initial velocity v_0, pulled down by a constant gravitational acceleration g:

$$s(t) \;=\; s_0 + v_0 t - \tfrac{1}{2} g t^2.$$

To justify this equation, note that the initial height is indeed $s(0) = s_0 + v_0(0) - \frac{1}{2}g(0^2) = s_0$. Also, s_0, v_0, g are constants, so:

$$v(t) = s'(t) = (s_0)' + (v_0 t)' - (\tfrac{1}{2}gt^2)' = v_0 - gt,$$

and the initial velocity is $v(0) = v_0$. The acceleration is $a(t) = v'(t) = -g$, which is the desired constant in the correct (downward) direction. Finally, the jerk is $j(t) = a'(t) = 0$, which is correct because gravity pulls steadily and never jerks.

EXAMPLE: Given standard gravity of 32 ft/sec^2 and initial height $s_0 = 5$ ft, how fast to throw a ball upward so that it stays airborne for 4 sec? The equation becomes $s(t) = 5 + v_0 t - 16t^2$, with the throw velocity v_0 an unknown constant. Landing at 4 sec means $s(4) = 0$, that is $5 + v_0(4) - 16(4^2) = 0$, and we can solve for $v_0 = 62.75$ ft/sec. (This is $62.75/1.47 \cong 43$ mph, which would require a strong arm.)

How high will the ball go from such a throw? The velocity is:

$$v(t) = (5 + 62.75t - 16t^2)' \;=\; 62.75 - 32t.$$

At the instant $t = t_1$ when the ball reaches the top of its arc, its velocity is zero. That is: $v(t_1) = 62.75 - 32t_1 = 0$ and $t_1 \cong 1.96$ sec. (This is not quite half the 4 sec interval, because the ball started out at $s_0 = 5$ ft.) The height at this instant is $s(1.96) \cong 66.5$ ft.

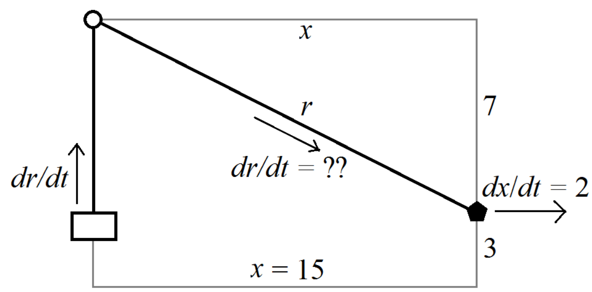
For $t < t_1$, the height $s(t)$ increases, the velocity $v(t) = 79 - 32t$ is positive, the ball moves upward; for $t > t_1$, $s(t)$ decreases, $v(t)$ is negative, the ball moves downward.

Note that the graph $s = 5 + 62.75t - 16t^2$ is a downward-curving parabola, but this is not the trajectory of the ball, since this model ignores horizontal motion: the ball might be going straight up and down.

Pulley example. Consider a weight hanging from a rope which stretches up to a pulley 10 ft above the floor, then to your hand, which is 3 ft above the floor and 15 ft horizontally from the pulley. If you walk away from the pulley at 2 ft/sec, how fast will the weight rise?

We want to find an unknown rate of change from a known rate which is related to it geometrically. To start any such problem, we draw a picture and label constant parts with their values: the lengths 3 and 7 below, which will not change as your hand moves horizontally. We label variable parts with letter names: the variable $h = h(t)$ is the horizontal distance from weight to hand, and $r = r(t)$ is the length of rope from pulley to hand, both functions of time t.

The problem specifies the current values of some variables, usually meaning at time $t = 0$: here $h(0) = 15$. Finally, for each variable we draw an arrow marked with its current rate of change: we know $h'(0) = 2$, and $r'(0)$ is the target rate which we aim to compute, since the weight goes upward at the same rate as r increases.



Next, we write equations implied by the geometry of the picture: the Pythagorean Theorem implies $r^2 = h^2 + 7^2$. To determine $r'(0)$, we compute $r(t)$ explicitly, and differentiate:

$$\begin{aligned} r(t) &= \sqrt{h(t)^2+49} \\ r'(t) &= \tfrac{1}{2}(h(t)^2+49)^{-1/2} \cdot \big(h(t)^2+49\big)' \\ &= \tfrac{1}{2}(h(t)^2+49)^{-1/2} \cdot 2h(t)h'(t). \\ &= \frac{h(t)h'(t)}{\sqrt{h(t)^2+49}}\,. \end{aligned}$$

Plugging in the current values at $t = 0$:

$$r'(0) = \frac{h(0)h'(0)}{\sqrt{h(0)^2+49}} = \frac{(15)(2)}{\sqrt{15^2+49}} = \frac{30}{\sqrt{274}} \cong 1.8 \text{ ft/sec}\,.$$

Notes by Peter Magyar `magyar@math.msu.edu`

We could do this a bit more simply by implicitly differentiating both sides of the equation $r^2 = h^2 + 7^2$, then solving for $r'(t)$:

$$\begin{aligned} \big(r(t)^2\big)' &= \big(h(t)^2 + 49\big)' \\ 2r(t)r'(t) &= 2h(t)h'(t) \\ r'(t) &= \frac{h(t)h'(t)}{r(t)}\,. \end{aligned}$$

Now, $r(0) = \sqrt{h(0)^2 + 49} = \sqrt{274}$, so plugging in current values: $r'(0) = \frac{(15)(2)}{\sqrt{274}}$ as before.

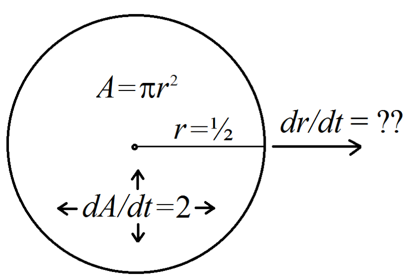
Warning: It is essential to plug in the current values only in the *last step*: if we substituted before differentiating, we would get: $(r(0))' = (\sqrt{h(0)^2 + 49})' = 0$ since the derivative of any constant (even a complicated constant) is *zero*.

Method for related rates problems

1. Draw a picture labeled with:
   - numerical constant values
   - letter variables and their known current values
   - arrows showing known current rates of change (derivatives)
   - an arrow for the unknown rate of change which is desired (the target rate)
2. Write an equation relating the variables according to the geometry of the picture.
3. Assuming each variable is a function of time t, take the derivative $\frac{d}{dt}$ of both sides of the equation, with the Chain Rule producing derivatives of the variables. If necessary, solve the derivative equation for the derivative which is desired.
4. Plug in the current values of the variables and rates to compute the target rate.

Ice block example. We saw a related rates problem in Notes §2.3, last page.

Spill radius example. A stream of water is spreading a circular puddle on the floor. If the puddle is 1 meter across, and the stream increases the area at a rate of 2 sq m/min, then how quickly is the puddle widening?



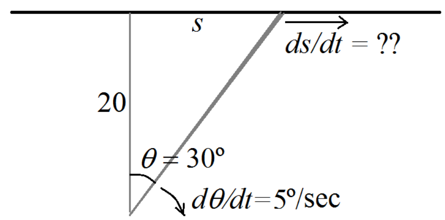
The variable quantities are the radius r and the area A. We know the current value $r(0) = \frac{1}{2}$ and the current rate $A'(0) = \frac{dA}{dt}|_{t=0} = 2$. The unknown rate which we must find is $r'(0)$. The area is related to the radius by the equation: $A = \pi r^2$. Differentiating the equation:

$$A'(t) \;=\; \pi\big(r(t)^2\big)' \;=\; 2\pi\, r(t)\, r'(t).$$

Solving for the target rate: $r'(t) = \frac{A'(t)}{2\pi\, r(t)}$, and $r'(0) = \frac{A'(0)}{2\pi\, r(0)} = \frac{2}{2\pi(\frac{1}{2})} = \frac{2}{\pi} \cong 0.64$ m/min.

It is important to check a real-world result for plausibility. The puddle's radius is growing (positive derivative) at a rate of about half a meter per minute, which is reasonable.

Searchlight example: A searchlight is shining along a wall 20 meters away. If the position of the light is 30° away from looking directly at the wall, and the light is turning at 5° per second, then what is the speed of the spotlight image moving along the wall?



The distance from the wall is the constant 20; the variable quantities are θ and s. The angle $\theta(t)$ has current value $\theta(0) = 30^\circ$ and current rate $\theta'(0) = 5^\circ/\mathrm{sec}$, and we seek to compute the unknown rate $s'(0) = \frac{ds}{dt}|_{t=0}$. From the definition of tangent, we have the equation: $\tan(\theta) = \frac{s}{20}$, so we can easily solve for $s = 20\tan(\theta)$. Differentiating (in Leibnitz notation this time):

$$\frac{ds}{dt} \;=\; \frac{d}{dt}(20\tan(\theta)) \;=\; 20\sec^2(\theta)\cdot\frac{d\theta}{dt},$$

since $\frac{d}{dx}\tan(x) = \sec^2(x)$ from the table in Notes §2.4. We do not need to solve for $\frac{ds}{dt}$, since we already solved for s before differentiating.

Finally, to plug in the current values of the angles, we must convert them to radians, because the trig differentiation formulas are *only valid for radian measure* (see last page of Notes §2.5). Thus:

$$\theta(0) \;=\; 30^\circ \;=\; 30(\tfrac{2\pi}{360}) \;=\; \tfrac{\pi}{6}\ \mathrm{rad},$$

$$\theta'(0) \;=\; \tfrac{d\theta}{dt}|_{t=0} \;=\; 5^\circ/\mathrm{sec} \;=\; 5(\tfrac{2\pi}{360}) \;=\; \tfrac{\pi}{36}\ \mathrm{rad/sec},$$

so the current speed is:

$$s'(0) \;=\; \frac{ds}{dt}\Big|_{t=0} \;=\; 20\sec^2(\tfrac{\pi}{6})\cdot\tfrac{\pi}{36} \;=\; \tfrac{20}{27}\pi \;\cong\; 2.3\mathrm{m/sec}.$$

Note that plugging in $\frac{d\theta}{dt}|_{t=0} = 5$ deg/sec instead of $\frac{\pi}{36} \cong 0.09$ rad/sec would give a wildly incorrect answer for $s'(0)$ in m/sec: the conversion of θ to radians is essential.

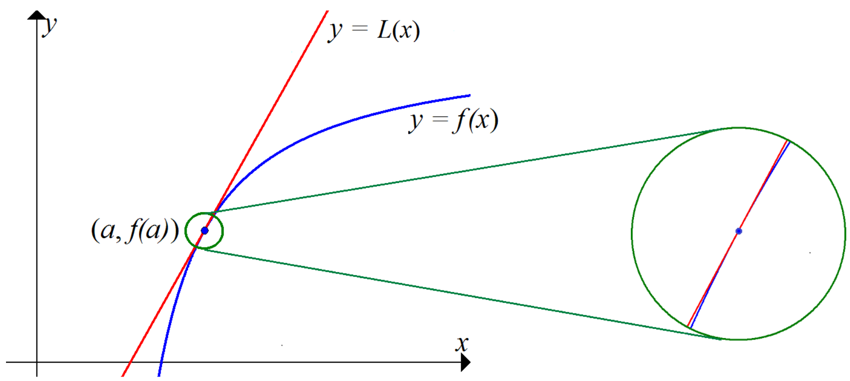
One last point: the problem specifies only the speed of θ, not the velocity toward or away from the wall, so we only know $\theta'(0) = \pm\frac{\pi}{36}$, either plus or minus, though in the picture we assumed it was plus. Thus we can only compute $s'(0) = \pm\frac{20}{27}\pi$, but in any case the speed is $|s'(0)| = \frac{20}{27}\pi$.

Tangent linear function. The geometric meaning of the derivative $f'(a)$ is the slope of the tangent to the curve $y = f(x)$ at the point $(a, f(a))$. The tangent line is itself the graph of a linear function $y = L(x)$, where:

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

This is correct because the line $y = f(a) + f'(a)(x-a)$ has slope $m = f'(a)$, and $L(a) = f(a) + f'(a)(a-a) = f(a)$, so the line passes through the point $(a, L(a)) = (a, f(a))$.

The value $f'(a)$ is not just the slope of the tangent line: it is also the slope of the graph itself, because as we zoom in toward $(a, f(a))$, the graph and the tangent line become indistinguishable*:



This suggests a further numerical meaning of the derivative: any function $f(x)$ is very close to being a linear function near a differentiable point $x = a$, so that $L(x)$ is a good approximation for $f(x)$ when x is close to a:

$$f(x) \;\approx\; L(x) \;=\; f(a) + f'(a)(x-a) \quad \text{for } x \approx a.$$

Much later in §11.10 of Calculus II, we will study Taylor series, which give much better, higher-order approximations to $f(x)$.

EXAMPLE: Find a quick approximation for $\sqrt{1.1}$ without a calculator. Clearly, this is close to $\sqrt{1} = 1$, but we want more accuracy. Take $f(x) = \sqrt{x}$, so $f'(x) = \frac{1}{2}x^{-1/2}$ and $f'(1) = \frac{1}{2}$. For x near $a = 1$, we have the linear function:

$$L(x) \;=\; f(1) + f'(1)(x-1) \;=\; 1 + \tfrac{1}{2}(x-1),$$

and the linear approximation:

$$\sqrt{1.1} \;=\; f(1.1) \;\cong\; L(1.1) = 1 + \tfrac{1}{2}(0.1) = 1.05.$$

A calculator gives: $\sqrt{1.1} \approx 1.049$, so our answer is correct to 2 decimal places with very little work. Furthermore, we get approximations for all other square roots near 1 for free, for example $\sqrt{0.96} \cong 1 + \frac{1}{2}(0.96-1) = 1-0.02 = 0.98$.

Notes by Peter Magyar `magyar@math.msu.edu`

*By contrast, if we zoom in toward a non-differentiable point, such as $(0, 0)$ for the graph $y = |x|$, the graph does *not* look more and more linear, but rather keeps its angular appearance.

EXAMPLE: Approximate $\sin(42°)$ without a scientific calculator. This is clearly close to $\sin(45°) = \frac{\sqrt{2}}{2} \approx 0.71$, so let us take $a = 45°$. Now, to use calculus with trig functions, we must always convert to radians: $a = 45(\frac{2\pi}{360}) = \frac{\pi}{4}$ rad. Thus $f'(a) = \sin'(\frac{\pi}{4}) = \cos(\frac{\pi}{4}) = \frac{\sqrt{2}}{2}$, and we have the linear function:

$$L(x) \;=\; \tfrac{\sqrt{2}}{2} + \tfrac{\sqrt{2}}{2}(x-\tfrac{\pi}{4}).$$

The linear approximation is:

$$\sin(42°) \;=\; \sin\!\big(42(\tfrac{2\pi}{360})\big) \;\approx\; L(x) \;=\; \tfrac{\sqrt{2}}{2} + \tfrac{\sqrt{2}}{2}\big(42(\tfrac{2\pi}{360})-\tfrac{\pi}{4}\big) \;=\; \tfrac{\sqrt{2}}{2} - \tfrac{\sqrt{2}}{2}\big(\tfrac{\pi}{60}\big) \;\approx\; 0.67.$$

A scientific calculator gives $\sin(42°) \approx 0.669$, so again the linear approximation is accurate to two decimal places.

Error sensitivity. We can rewrite the linear approximation $f(x) \approx f(a) + f'(a)(x-a)$ as:

$$\Delta f \;=\; f(x)-f(a) \;\approx\; f'(a)(x-a) \;=\; f'(a)\,\Delta x.$$

That is, we can approximate the change in $f(x)$ away from $f(a)$, in proportion to the change in x away from a. In Leibniz notation, with $y = f(x)$, we write this as:

$$\Delta y \;\approx\; \tfrac{dy}{dx}\,\Delta x.$$

Here we mean $\frac{dy}{dx} = \frac{dy}{dx}|_{x=a} = f'(a)$. If we think of Δx as an error from an intended input value $x = a$, then $\Delta f \approx f'(a)\,\Delta x$ approximates the error from the intended output $f(a)$.

EXAMPLE: A disk of radius $r = 5$ cm is to be cut from a metal sheet weighing 3 g/cm^2. If the radius is measured to within an error of $\Delta r = \pm 0.2$ cm, what is the approximate range of error in the weight? This is the kind of error-control problem from our limit analyses in Notes §1.7, only now we have the powerful tools of calculus to give a simple answer.

The weight is the density 3 times the area πr^2, given by the function:

$$W = W(r) = 3\pi r^2 \qquad \text{with} \qquad W(5) = 75\pi \approx 235.6,$$

and we aim to find the error ΔW away from this intended value. Since:

$$\tfrac{dW}{dr} \;=\; 3\pi(2r) \;=\; 6\pi r \qquad \text{and} \qquad \tfrac{dW}{dr}|_{r=5} = 30\pi,$$

we have the approximate error:

$$\Delta W \;\approx\; \tfrac{dW}{dr}\Delta r \;=\; 30\pi\,\Delta r.$$

Thus, for $\Delta r = \pm 0.2$, we have $\Delta W \approx 30\pi(0.2) \approx 18.8$. That is:

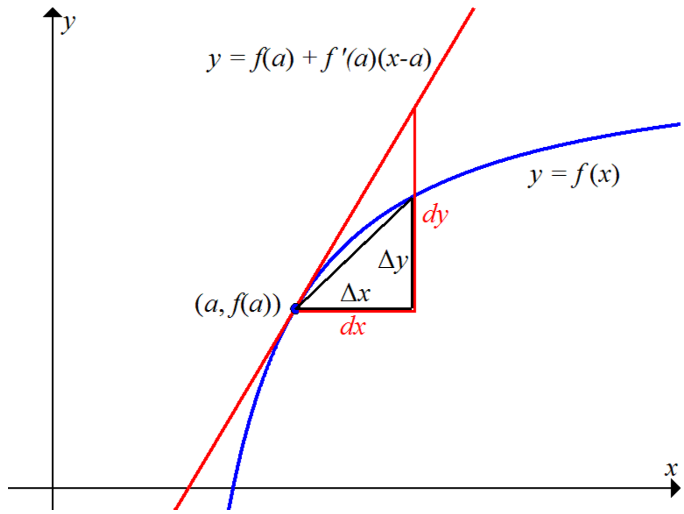
$$r = 5 \pm 0.2 \text{ cm} \qquad \Longrightarrow \qquad W \;\approx\; 235.6 \pm 18.8 \text{ g}\,.$$

The point here is not just the specific error estimate, but the formula which gives, for any small input error Δr, the resulting output error $\Delta W \approx 30\pi\,\Delta r \approx 94\,\Delta r$. The coefficient 30π measures the *sensitivity* of the output W to an error in the input r.

Differential notation. For $y = f(x)$, we rewrite a small Δx as dx, and we define:

$$dy \;=\; \tfrac{dy}{dx}\,dx \qquad \text{and} \qquad df \;=\; f'(x)\,dx.$$

The dependent quantity dy is called a differential: we can think of it as the linear approximation to Δy, as pictured below:



EXAMPLE: We can rewrite the approximation in the previous example as:

$$\Delta W \ \approx\ dW \ =\ \tfrac{dW}{dr}\,dr \ =\ \tfrac{d}{dr}(3\pi r^2)\,dr \ =\ 6\pi r\,dr.$$

Here dr is just another notation for Δr, and the approximation $\Delta W \approx 6\pi r\,\Delta r$ is valid near any particular value of r, such as $r = 5$ in the example.

Linear Approximation Theorem. How close is the approximation $\Delta y \approx dy$, or equivalently $f(x) \approx L(x) = f(a) + f'(a)(x{-}a)$? In fact, the difference between $f(x)$ and $L(x)$ is not only small compared to $\Delta x = x{-}a$, but usually proportional to $(\Delta x)^2 = (x{-}a)^2$, which becomes tiny as $\Delta x \to 0$. (E.g. if $\Delta x = 0.01$, then $(\Delta x)^2 = 0.0001$.)

Also, the slower the derivative $f'(x)$ changes near $x = a$, the closer $y = f(x)$ is to its linear approximation, and this is measured by the rate of change of $f'(x)$, namely the second derivative $f''(x)$. The following theorem gives an upper bound on the error in the linear approximation, $\varepsilon(x) = f(x) - L(x)$.

Theorem: Suppose $f(x)$ is a function such that $|f''(x)| < B$ on the interval $x \in [a{-}\delta, a{+}\delta]$. Then, for all $x \in [a{-}\delta, a{+}\delta]$, we have:

$$f(x) \ =\ f(a) + f'(a)(x{-}a) + \varepsilon(x), \qquad \text{where} \qquad |\varepsilon(x)| < B|x{-}a|^2.$$

We give the proof, based on the Mean Value Theorem, later in §3.2.

EXAMPLE: For $f(x) = \sqrt{x}$ near $x = 1$, we have $f'(x) = \frac{1}{2}x^{-1/2}$ and $f'(1) = \frac{1}{2}$. Also $f''(x) = -\frac{1}{4}x^{-3/2}$, and on the interval $x \in [0.9, 1.1]$, we have:

$$|f''(x)| \le |f''(0.9)| = \tfrac{1}{4}(0.9)^{-3/2} \approx 0.29 < \tfrac{1}{3}.$$

Thus we may take $B = \frac{1}{3}$, and find that:

$$\sqrt{x} \ =\ \sqrt{1} + \tfrac{1}{3}(x{-}1) + \varepsilon(x), \qquad \text{where} \qquad |\varepsilon(x)| < \tfrac{1}{3}|x{-}1|^2.$$

For example, the error at $x = 1.1$ is $|\varepsilon(1.1)| < \frac{1}{3}(0.1)^2 < 0.004$, so:

$$\sqrt{1.1} \ =\ 1 + \tfrac{1}{2}(0.1) \pm 0.004 \ =\ 1.05 \pm 0.004\,.$$

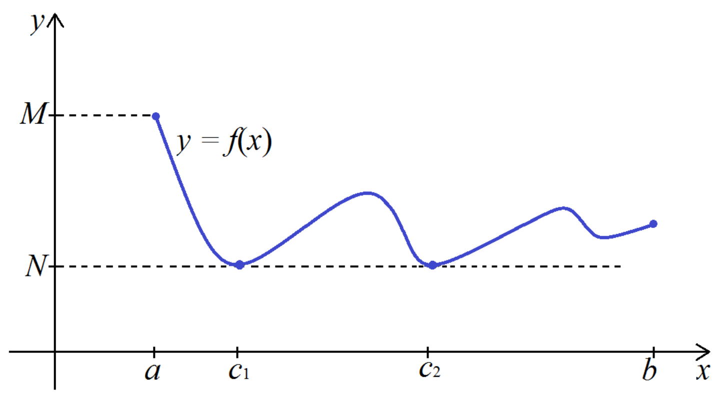
Extreme Values

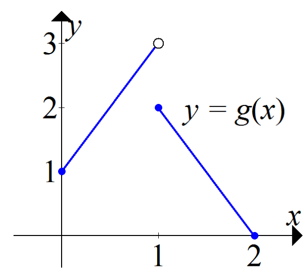
Absolute maxima and minima. In many practical problems, we must find the largest or smallest possible value of a function over a given interval.

> *Definition:* For a function $f(x)$ defined on an interval $x \in [a, b]$, an *absolute maximum* (or *global maximum*) is a point $c \in [a, b]$ such that $f(c) \geq f(x)$ for all $x \in [a, b]$. That is, $f(c)$ is the largest output value of the function at any input point in its domain. We say $x = c$ is a *maximum point* and $f(c)$ is the *maximum value*.

We define an *absolute minimum* similarly, and both maximums and and minimums are *extremums** or *extreme points*. Note that the maximum value M (the largest possible output) is unique, but $f(x)$ could touch this value at several input points $c_1, c_2, \ldots \in [a, b]$, all having $f(c_1) = f(c_2) = \cdots = M$.

EXAMPLE: At left below, the function $y = f(x)$ on the interval $[a, b]$ has one absolute maximum point, the left endpoint $x = a$ with $f(a) = M$, so that $(a, f(a))$ is the highest point on the graph; and it has two absolute minimum points $x = c_1, c_2$ with $f(c_1) = f(c_2) = N$, so that $(c_1, f(c_1))$ and $(c_2, f(c_2))$ are the lowest points on the graph.





> *Extremal Value Theorem:* If $f(x)$ is continuous on the closed, finite interval $x \in [a, b]$, then $f(x)$ possesses at least one maximum point and one minimum point.

A proof valid for all possible continuous functions would require sophisticated Real Analysis concepts as in Math 320. To see that the theorem is not obvious, consider the function $y = g(x)$ graphed at right above. It is not continuous because the graph has a break, so the Theorem does not guarantee an absolute maximum; and indeed there is *no absolute maximum*. Instead, the function approaches $y = 3$ as $x \to 1^-$ (i.e. $x = 1-\delta$ for small $\delta > 0$), but it never actually reaches $y = 3$ because it suddenly drops to $g(1) = 2$. Thus, for any given output $g(c)$, we can find some slightly larger output $g(1-\delta) > g(c)$ for a tiny $\delta > 0$, so no $g(c)$ is largest. However, there is an absolute min point $x = 2$.

Notes by Peter Magyar `magyar@math.msu.edu`

*The Latin plurals of *maximum, minimum, extremum* are *maxima, minima, extrema.*

Local maxima and minima. A broader, but still useful, concept is that of a *local extremum*: this is a point where the graph has a hill or valley, but not necessarily the highest or lowest one.

> *Definition:* For a function $f(x)$ defined on an interval $x \in [a,b]$, a *local maximum* (or *relative maximum*) is a point $c \in [a,b]$ such that $f(c)$ is the largest output value for any input point *nearby* $x = c$.
>
> Formally, there is a small $\delta > 0$ such that $f(c) \geq f(x)$ for all $x \in [c-\delta, c+\delta]$; or $x \in [a, a+\delta]$ if $c = a$; or $x \in [b-\delta, b]$ if $c = b$.

Clearly, an absolute maximum must also be a local maximum. To illustrate, the function $f(x)$ in the figure at left above has four local maximum points, the two endpoints and the two hill tops; and it has three local minimum points, all valley bottoms. (The discontinuous $g(x)$ does not have any local maxima.)

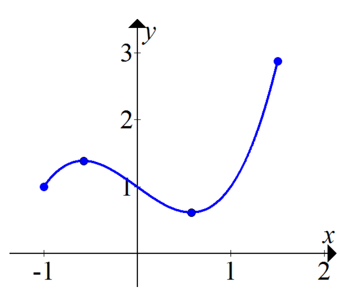
Vanishing derivatives. Calculus makes finding extremums surprisingly easy. We have already seen a real-world example at the end of Notes §2.7, where we asked for the maximum height of a ball whose height at time t is given by the ballistic function $s(t) = 5 + 79t - 16t^2$. The ball is highest at the moment when it passes from rising to falling and its velocity is zero: $t = c$ with $v(c) = 0$, where $v(t) = s'(t) = 79 - 32t$; and we solve to get $c = \frac{79}{32}$. That is, if $t = c$ is the maximum point of $s(t)$, then $s'(c) = 0$.

This also makes sense graphically. If $x = c$ is a local maximum of a function $f(x)$, then $(c, f(c))$ is most likely a hill-top of the graph $y = f(x)$, and the tangent line is horizontal at this point, having zero slope. But the tangent slope is the derivative $f'(c)$, so if $t = c$ is a local maximum, then $f'(c) = 0$. The same goes for a local minimum or valley-bottom.

> *First Derivative Theorem:* If $f(x)$ has a local maximum or minimum at $x = c$, which is not an endpoint of the interval of definition, and $f(x)$ is differentiable at this point, then $f'(c) = 0$.

Proof: Suppose $f(x)$ has a local minimum, so $f(x) \geq f(c)$ for $c-\delta \leq x \leq c+\delta$. For $c < x \leq c+\delta$, we have $\frac{f(x)-f(c)}{x-c} = \frac{+}{+} \geq 0$, so $f'(c) = \lim_{x\to c^+} \frac{f(x)-f(c)}{x-c} \geq 0$. For $c-\delta \leq x < c$, we have $\frac{f(x)-f(c)}{x-c} = \frac{+}{-} \leq 0$, so $f'(c) = \lim_{x\to c^-} \frac{f(x)-f(c)}{x-c} \leq 0$. Thus $f'(c)$ is both positive and negative, which can only mean $f'(c) = 0$.

EXAMPLE: We wish to find the maxima and minima, both local and absolute, of $f(x) = x^3 - x + 1$ on the interval $x \in [-1, \frac{3}{2}]$. Since $f(x)$ is continuous (by the Limit Laws), the Extremal Value Theorem guarantees there is at least one of each type of point.



Exactly where are the hill-top and the valley-bottom points? Since $f(x)$ is differentiable at every point, the First Derivative Theorem means that all local maximum and minimum points must be endpoints or solutions of $f'(x) = 0$, namely $3x^2 - 1 = 0$, or $x = \pm\frac{1}{\sqrt{3}} \approx \pm 0.58$. The graph shows that the local maxima are the hill-top $x = -\frac{1}{\sqrt{3}}$ and the right endpoint $x = \frac{3}{2}$, and the one with the larger output is the absolute maximum: $f(-\frac{1}{\sqrt{3}}) \approx 1.4 < f(\frac{3}{2}) \approx 2.9$, so the endpoint $x = \frac{3}{2}$ is the absolute maximum point. Similarly, the local minima are $x = -1$ and $x = \frac{1}{\sqrt{3}}$ with $f(-1) = 1 > f(\frac{1}{\sqrt{3}}) \approx 0.61$, so $x = \frac{1}{\sqrt{3}}$ has the smaller output and is the absolute minimum point.

Critical points. The above example illustrates the method for identifying all relevant candidates for the absolute maximum and minimum: the endpoints and the points where the derivative vanishes, and also possibly where the derivative is not defined because the graph has a corner or a discontinuity.

> *Definition:* For a function $f(x)$, a *critical point* (or *critical number*) is a point $x = c$ where the derivative is either zero or the function is not differentiable: $f'(c) = 0$ or undefined.

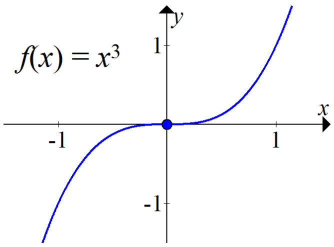
Method for absolute maxima and minima problems.

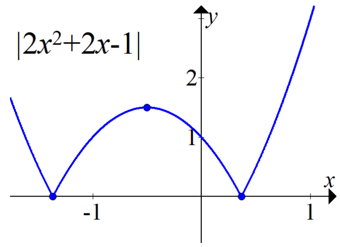
1. Given $f(x)$ on an interval $x \in [a, b]$, determine the critcal points (critical numbers) $x = c$ such that $f'(c) = 0$ or undefined. Be sure to consider only those $c \in [a, b]$, discarding any critical points outside the relevant interval.

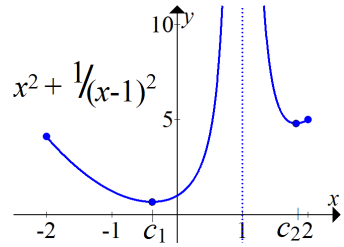
2. If $f(x)$ is continuous, find $f(x)$ for all critical points $x = c$ and for the endpoints $x = a, b$. Those points with the largest output are the absolute maximum points, and those with smallest values are the absolute minima.

3. If $f(x)$ has any discontinuity $x = c$, examine nearby $x \to c^+$ and $x \to c^-$ to see if the outputs $f(x)$ become larger or smaller than in Step 2.

Most functions are continuous and differentiable as in the previous example, and it is enough to perform Step 1 with $f'(c) = 0$, then Step 2. Below we illustrate some more complicated situations.







EXAMPLE: Not every critical point must be a local maximum or minimum. For $f(x) = x^3$, solving $f'(x) = 3x^2 = 0$ gives $x = 0$ as the unique critical point. The graph (above left) has a horizontal slope which is neither a hill-top nor a valley-bottom, but rather a *stationary point*, where the function pauses in its rise. This does not derail the Method, since it only gives an extra candidate for the absolute max/min, which will be discarded because its output value is neither largest nor smallest over any given interval.

EXAMPLE: Let $f(x) = |2x^2 + 2x - 1|$, with graph at center above. Recall that $\frac{d}{dx}|x| = \text{sgn}(x) = \frac{|x|}{x}$, which is undefined when $x = 0$. By the Chain Rule:

$$f'(x) \;=\; \text{sgn}(2x^2{+}2x{-}1) \cdot (2x^2{+}2x{-}\,1)' \;=\; \text{sgn}(2x^2{+}2x{-}1) \cdot (4x{+}2).$$

Since sgn( ) is never zero, we have $f'(x) = 0$ when the second factor vanishes: $4x + 2 = 0$, or $x = \frac{1}{2}$.

But this is not the only critical point, since we must also consider when $f'(x)$ is undefined. This happens when the first factor $\text{sgn}(2x^2 + 2x - 1)$ is undefined, namely when $2x^2 + 2x - 1 = 0$, or $x = \frac{1}{4}(-2\pm\sqrt{12})$ by the Quadratic Formula. These are the corners of the graph sitting on the x-axis: we must not skip them, since they are actually the absolute minimum points.

EXAMPLE: Let $f(x) = x^2 + \frac{1}{(x-1)^2}$ on the interval $x \in [-2, 2]$, graph above right, with:

$$f'(x) \;=\; \big(x^2\big)' + \big((x{-}1)^{-2}\big)' \;=\; 2x + (-2)(x{-}1)^{-3}(x{-}2)' \;=\; \frac{2(x^4{-}3x^3{+}3x^2{-}x{-}1)}{(x{-}1)^3}\,.$$

We have $f'(x) = 0$ when the numerator vanishes, $x^4{-}3x^3{+}3x^2{-}x{-}1 = 0$, and graphing this degree 4 polynomial gives approximate solutions $x = c_1 \approx -0.38$ and $c_2 \approx 1.82$ with $f(c_1) \approx 0.67$ and $f(c_2) \approx 4.80$. The endpoints $x = \pm 2$ give $f(-2) = \frac{37}{9} \approx 4.11$ and $f(2) = 5$. We might be tempted to take the largest of these outputs as the absolute maximum, but clearly none of these is the highest point of the graph.

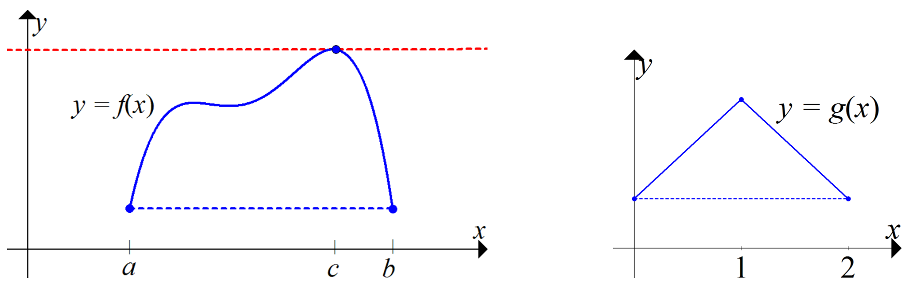
We neglected to consider when $f'(x)$ is undefined: this is when the denominator $(x{-}1)^2 = 0$, or $x = 1$. This is a discontinuity, so by Step 3 we must consider not only $f(1)$, which is undefined, but also a small interval around $x = 1$. In fact, we have a vertical asymptote, and $\lim_{x\to 1^-} f(x) = \lim_{x\to 1^+} f(x) = \infty$.

That is, $f(x)$ can get as large as desired for x close enough to 1. There is *no absolute maximum.* However, the rising asymptotes do not affect the absolute minumum, which is still the smallest of the outputs at the other critical points, namely $f(c_1) \approx 0.67$. Note that since $f(x)$ is not continuous, the Extremal Value Theorem does not guanrantee an absolute max or min; and in fact the max does not exist, but the min does.

Vanishing derivatives. We will prove some basic theorems which relate the derivative of a function with the values of the function, culminating in the Uniqueness Theorem at the end. The first result is:

> *Rolle's Theorem:* If $f(x)$ is continuous on a closed interval $x \in [a, b]$ and differentiable on the open interval $x \in (a, b)$, and $f(a) = f(b)$, then there is some point $c \in (a, b)$ with $f'(c) = 0$.

Here $x \in [a, b]$ means $a \le x \le b$, and $x \in (a, b)$ means $a < x < b$. See the graph at left for an example: no matter how the curve wiggles, it must be horizontal somewhere.



Physically, imagine $f(t)$ represents the height of a rocket at time t, starting and finishing on its launch pad over the time interval $t \in [a, b]$. The theorem says there must be a pause in the motion where $f'(t) = 0$: this is the moment the rocket runs out of fuel and starts to fall.

Proof of Theorem. Assume $f(x)$ satisfies the hypotheses* of the Theorem. The Extremal Value Theorem (§3.1) guarantees that the continuous function $f(x)$ has at least one absolute maximum point $x = c_1 \in [a, b]$.

- If $c_1 \ne a, b$, then $c_1 \in (a, b)$, and the First Derivative Theorem (§3.1) says that $f'(c_1) = 0$.
- On the other hand, if $c_1 = a$ or b, then $f(c_1) = f(a) = f(b)$. Still, $f(x)$ also has an absolute minimum point $x = c_2$. If $c_2 \in (a, b)$, then $f'(c_2) = 0$ as before.
- The only case left is if $c_1 = a$ or b, and also $c_2 = a$ or b, so that $f(c_1) = f(c_2) = f(a) = f(b)$. Since the maximum and minimum values are the same, $f(x)$ cannot move above or below $f(a)$. Thus, $f(x)$ can only be a constant function, and $f'(c) = 0$ for *all* $c \in (a, b)$.

In every case, the conclusion† holds, Q.E.D.‡

Notes by Peter Magyar `magyar@math.msu.edu`

*In formal mathematics, *hypothesis* (plural *hypotheses*) means the "if" part of a theorem, the setup which is given or assumed. In our theorem, the three hypotheses are: $f(x)$ is continuous on $[a, b]$, $f(x)$ is differentiable on (a, b), and $f(a) = f(b)$.

†*Conclusion* means the "then" part of a theorem, the payoff which is to be deduced from the hypothesis: in our theorem, that $f'(c) = 0$.

‡Latin *quod erat demonstrandum*, "which was to be shown", the traditional end of a proof.

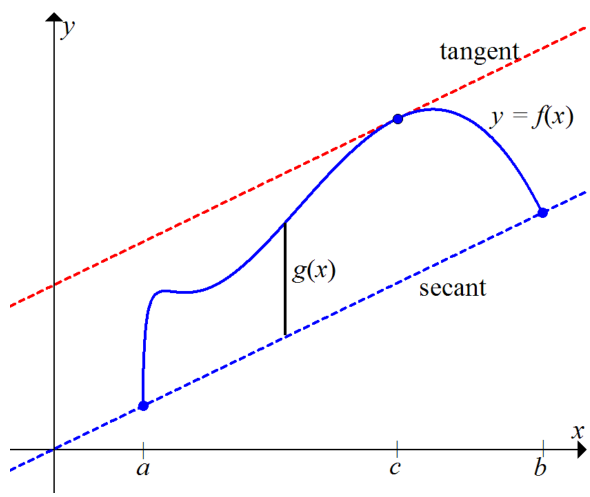
For Rolle's Theorem, as for most well-stated theorems, all the hypotheses are necessary to be sure of the conclusion. In the graph at right above, $y = g(x)$ has a corner and $g'(1)$ does not exist, so just one hypothesis fails at just one point. But already the conclusion is false: $g'(c) = 1$ for $c < 1$ and $g'(c) = -1$ for $c > 1$, but nowhere is $g'(c) = 0$. In physical terms, the velocity jumps instantaneously from 1 to -1 like an idealized ping-pong ball, and there is no well-defined velocity at the moment of impact.

Derivatives versus difference quotients. Throughout our theory, the derivative $f'(a)$ has been shadowed by the difference quotient $\frac{\Delta f}{\Delta x} = \frac{f(b)-f(a)}{b-a}$, over some interval $[a, b]$. Numerically, the difference quotient is an approximation to the derivative: $\frac{df}{dx} \approx \frac{\Delta f}{\Delta x}$, provided Δx is small. In physical terms, the difference quotient is the average rate of change of $f(x)$ over $x \in [a, b]$. Geometrically in terms of the graph $y = f(x)$, the difference quotient is the slope of the secant line cutting through the points $(a, f(a))$ and $(b, f(b))$.

Now we come to the most powerful result of this section, which says that the derivative is sometimes exactly equal to the difference quotient (the slope of one tangent is equal to the slope of the secant):

> *Mean Value Theorem (MVT):* If $f(x)$ is continuous on a closed interval $x \in [a, b]$ and differentiable on the open interval $x \in (a, b)$, then there is some point $c \in (a, b)$ with $f'(c) = \frac{f(b)-f(a)}{b-a}$.

See the picture below for an example: as the graph rises from $(a, f(a))$ to $(b, f(b))$, at some points the tangent line must be parallel to the secant line.



Note that Rolle's Theorem is the special case of MVT in which the secant line is horizontal. In fact, we will prove MVT for a general $f(x)$ by cooking up a new function $g(x)$ for which Rolle's Theorem applies, then translating Rolle's conclusion back in terms of $f(x)$.

Proof of MVT. Suppose $f(x)$ satisfies the hypotheses. Then define a new function $g(x)$, shown in the picture, which measures the height from the graph $y = f(x)$ down to the secant line $y = f(a) + \frac{f(b)-f(a)}{b-a}(x-a)$:

$$g(x) \; = \; f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x-a).$$

Then $g(x)$ is continuous on $[a,b]$ by the Limit Laws (§1.6), and differentiable on (a,b) by the Derivative Rules (§2.3). In fact,

$$g'(x) \;=\; f'(x) - 0 - \tfrac{f(b)-f(a)}{b-a}(1-0) \;=\; f'(x) - \tfrac{f(b)-f(a)}{b-a},$$

since $f(a)$ and $\frac{f(b)-f(a)}{b-a}$ are constants (having no x in them).

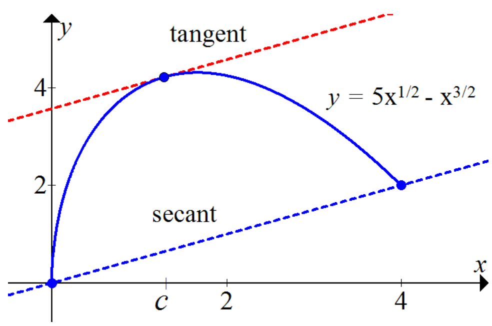
Also, we can easily compute that $g(a) = g(b) = 0$, so all the hypotheses of Rolle's Theorem hold for $g(x)$. Thus the conclusion of Rolle's Theorem also holds: there is some $c \in (a,b)$ with $g'(c) = 0$. That is,

$$g'(c) \;=\; f'(c) - \tfrac{f(b)-f(a)}{b-a} \;=\; 0,$$

which means $f'(c) = \frac{f(b)-f(a)}{b-a}$, Q.E.D.

The Mean Value Theorem does not give any way to find the particular $c \in (a,b)$ in the conclusion, so if we want this value in a particular case, we must solve for x in the equation $f'(x) = \frac{f(b)-f(a)}{b-a}$; however the Theorem will guarantee that there is some solution.

EXAMPLE: Let $f(x) = 5\sqrt{x} - x\sqrt{x}$ over the interval $[a,b] = [0,4]$.



To check the hypotheses of MVT, note that $\sqrt{x}$ is continuous for all $x \geq 0$, and thus over $[0,4]$. As for differentiability:

$$f'(x) \;=\; \left(5x^{1/2} - x^{3/2}\right)' \;=\; \tfrac{5}{2}x^{-1/2} - \tfrac{3}{2}x^{1/2}$$

is defined for $x > 0$, and hence over $x \in (0,4)$: the hypothesis allows $f'(a) = f'(0)$ to be undefined. Thus we conclude there must be some $c \in (0,4)$ with $f'(c) = \frac{f(b)-f(a)}{b-a} = \frac{2-0}{4-0} = \frac{1}{2}$. If we wish to find this c, we must solve:

$$f'(x) \;=\; \tfrac{5}{2}x^{-1/2} - \tfrac{3}{2}x^{1/2} \;=\; \tfrac{1}{2}\,,$$

which is equivalent to $3x + \sqrt{x} - 5 = 0$. Substituting the variable $u = \sqrt{x}$ gives $3u^2 + u - 5 = 0$, so the Quadratic Formula gives:

$$u \;=\; \sqrt{x} \;=\; \tfrac{-1\pm\sqrt{1^2-4(3)(-5)}}{2(3)} \;=\; \tfrac{-1\pm\sqrt{61}}{6}.$$

The negative solution is impossible, and the positive one gives $x = c = \left(\frac{\sqrt{61}-1}{6}\right)^2 \approx 1.29$, which agrees with the picture.

Mathematical and physical uniqueness. We come to the most important result of this section. Let $f(x)$ be a continuous function on an interval $x \in [p, q]$.

Uniqueness Theorem:

(a) If $f'(x) = 0$ for all $x \in (p, q)$, then $f(x) = C$, a constant function.

(b) If $f(x), g(x)$ have the same derivative $f'(x) = g'(x)$ for all $x \in (p, q)$, then $f(x) = g(x) + C$ for some constant C.

(c) If $f(x), g(x)$ have the same derivative $f'(x) = g'(x)$ for all $x \in (p, q)$, and the same initial value $f(c) = g(c)$ for some $c \in [p, q]$, then $f(x) = g(x)$.

Proof. (a) Assume the hypothesis $f'(x) = 0$ for all $x \in (p, q)$, and consider two outputs $f(a)$ and $f(b)$ for any $a < b$ in $[p, q]$. Applying the Mean Value Theorem to the smaller interval $[a, b]$, we get $\frac{f(b)-f(a)}{b-a} = f'(c) = 0$, since all derivatives $f'(x)$ are zero. Multiplying by $b-a$, we get $f(b)-f(a) = (b-a)\,0 = 0$, so $f(b) = f(a)$. That is, all the outputs of $f(x)$ are equal, and $f(x)$ is constant.

(b) Assume the hypothesis $f'(x) = g'(x)$ for all $x \in (a, b)$. Now the function $h(x) = f(x) - g(x)$ has $h'(x) = f'(x) - g'(x) = 0$, so we can apply part (a) to conclude that $h(x)$ is constant, $h(x) = f(x) - g(x) = C$, and $f(x) = g(x) + C$.

(c) In the situation of (b), we also assume $f(c) = g(c)$. By (b), we know $f(x) = g(x) + C$ for all x. In particular for $x = c$, we have $C = f(c) - g(c) = 0$, so $f(x) = g(x) + C = g(x)$, Q.E.D.

To see the significance of this theorem, recall from §2.7 the Ballistic Equation, which gives the height $s(t)$ of an object thrown straight up from initial height $s(0) = s_\circ$, with initial velocity $s'(0) = v_\circ$. The constant gravitational acceleration $-g$ means the velocity should steadily decrease with slope $-g$, so that $s'(t) = v_\circ - gt$. Now, the function

$$s(t) = s_\circ + v_\circ t - \tfrac{1}{2} g t^2$$

does indeed satisfy all these conditions.

But does this guarantee we have the correct function $s(t)$? What if there were some other function $\widetilde{s}(t)$ with the *same derivative* $\widetilde{s}'(t) = s'(t)$ and the *same initial value* $\widetilde{s}(0) = s(0)$? Then $\widetilde{s}(t)$ would be just as good a candidate to give the height of the object, and our mathematical theory would not produce a clear physical prediction. However, the Uniqueness Theorem (c) shows that $\widetilde{s}(t) = s(t)$: there is only one mathematical solution to the equation.

Experiment shows that objects launched in exactly the same way always fly the same way, not according to $s(t)$ in some experiments and a different $\widetilde{s}(t)$ in other experiments. This is what we mean by *physical law.* Our Theorem shows the mathematical solution has the same uniqueness as experimental results.[§]

[§]The theory of quantum mechanics, however, which explains atomic-scale phenomena, goes beyond the framework of deterministic laws, incorporating randomness not just as error (experiments are never perfectly controlled), but as an essential part of the setup. It requires a yet higher mathematical theory, in which we apply calculus not to specific positions of objects, but to probability distributions on all possible positions.

Derivative controls direction. We can similarly use the MVT to prove the expected ways that the sign of the derivative $f'(x)$ should control the increasing/decreasing behavior of the function $f(x)$. For example:

Theorem: A function with positive derivative is increasing. That is, if $f'(x) > 0$ for all x in an interval, then $f(a) < f(b)$ for any $a < b$ within the interval.

Proof: Assume $f'(x) > 0$ for all $x \in (a,b)$. By MVT, $f'(c) = \frac{f(b)-f(a)}{b-a}$ for some $c \in (a,b)$, so $f(b) - f(a) = (b-a)\,f'(c) > 0$ by assumption, and $f(b) > f(a)$.

Cauchy Mean Value Theorem. This is a generalized form of the MVT involving two functions $f(x)$ and $g(x)$, stating the quotient $\frac{\Delta f}{\Delta g} = \frac{f(b)-f(a)}{g(b)-g(a)}$ is equal to $\frac{f'(c)}{g'(c)}$ at some point c, provided the denominators are non-zero. Or in multiplied-out form:

Theorem. If $f(x), g(x)$ are continuous on $[a,b]$, differentiable on (a,b), then there is some $c \in (a,b)$ with

$$(f(b)-f(a))\,g'(c) \;=\; (g(b)-g(a))\,f'(c).$$

Proof. Apply Rolle's Theorem to $h(x) = (f(b)-f(a))g(x) - (g(b)-g(a))f(x)$, which has $h(a) = h(b)$. Then $h'(c) = 0$ gives the formula.

Proof of Linear Approximation Error Estimate. As a final application of MVT, consider as in §2.9 the linear approximation of a function $f(x)$ centered at $x = a$:

$$f(x) \;\approx\; L_a(x) \;=\; f(a) + f'(a)(x-a).$$

According to the Linear Approximation Theorem, the error is controlled by the second derivative: if $|f''(x)| \le B$ over an interval $x \in (a-\delta, a+\delta)$, then:

$$|f(x) - L_a(x)| \;\le \tfrac{1}{2}B\,|x-a|^2 \;\text{ for }\; x \in (a-\delta, a+\delta).$$

Proof. In our formulas, we consider x as a variable and a as an unspecified constant, but this is merely a point of view. We may instead hold x as a fixed value and allow a to vary, indicating this by replacing a with the variable t. Then we can take the derivative of the error function with respect to t, while keeping x constant:

$$\varepsilon(t) \;=\; f(x) - L_t(x) \;=\; f(x) - f(t) - f'(t)(x-t)$$

$$\varepsilon'(t) \;=\; 0 - f'(t) - \big(f'(t)(-1) + f''(t)(x-t)\big) \;=\; -f''(t)(x-t).$$

Now apply MVT to the interval $[a,x]$:[¶] there is some $t = c \in (a,x)$ with:

$$\frac{\varepsilon(x) - \varepsilon(a)}{x-a} \;=\; \varepsilon'(c).$$

Considering that $\varepsilon(x) = 0$, we find that:

$$\varepsilon(a) \;=\; -\varepsilon'(c)\,(x-a) \;=\; f''(c)\,(x-t)(x-a).$$

Finally, we use the hypothesis $|f''(c)| < B$ along with $|x-c| < |x-a|$ to get:

$$|f(x) - L_a(x)| = |\varepsilon(a)| < |f''(c)| \cdot |x-c| \cdot |x-a| < B\,|x-a|^2.$$

This is a slightly weaker upper bound than desired, since it is missing the factor of $\frac{1}{2}$.

To get the strong upper bound, we apply the Cauchy Mean Value Theorem to the functions $\varepsilon(t)$ and $g(x) = (x-t)^2$. There is some $c \in (a,x)$ with $\varepsilon'(c)/g'(c) = (\varepsilon(x) - \varepsilon(a))/(g(x) - g(a))$, so that $\varepsilon(a) = -\varepsilon'(c)(x-a)^2/2(x-c) = \frac{1}{2}f''(c)(x-a)^2$.

[¶] This assumes $a < x$, but the argument adapts easily to the other case $x < a$.

Increasing and decreasing functions. We will see how to determine the important features of a graph $y = f(x)$ from the derivatives $f'(x)$ and $f''(x)$, summarizing our Method on the last page. First, we consider where the graph is rising $\nearrow$ and falling $\searrow$. Formally:

> *Definition:* A function $f(x)$ is *increasing* on the interval $[a,b]$ whenever $f(x_1) < f(x_2)$ for every pair of inputs $x_1 < x_2$ in $[a,b]$; and $f(x)$ is *decreasing* on $[a,b]$ whenever $f(x_1) > f(x_2)$ for every $x_1 < x_2$.

We can determine this with derivatives: the graph rises where its slope is positive.

> *Increasing/Decreasing Theorem:* Let $f(x)$ be continuous on $[a,b]$.
> - If $f'(x) > 0$ for all $x \in (a,b)$,* then $f(x)$ is increasing on $[a,b]$.
> - If $f'(x) < 0$ for all $x \in (a,b)$, then $f(x)$ is decreasing on $[a,b]$.

Proof. Assume $f'(x) \geq 0$ for all $x \in (a,b)$, and consider any $x_1 < x_2$ in $[a,b]$. Applying the Mean Value Theorem to the interval $[x_1, x_2]$, we have:

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(c) > 0,$$

since all derivatives $f'(x)$ are positive. Multiplying by $x_2 - x_1 > 0$, we get:

$$f(x_2) - f(x_1) > (x_2 - x_1)0 = 0, \quad \text{so } f(x_2) > f(x_1).$$

That is, for all $x_1 < x_2$ in the interval $[a,b]$, we have $f(x_1) < f(x_2)$, and $f(x)$ is increasing. The second statement of the Theorem is proved similarly. Q.E.D.

EXAMPLE: For $f(x) = x^5 - 15x^3$, let us determine the rough shape of the graph by examining the derivative:

$$f'(x) = 5x^4 - 45x^2 = 5x^2(x^2-9) = 5x^2(x-3)(x+3).$$

Since $f'(x)$ is defined everywhere, the critical points (or critical numbers) are the solutions of $f'(x) = 0$, namely $x = -3, 0, 3$.

x		-3		0		3	
$f'(x)$	$+$	0	$-$	0	$-$	0	$+$
$f(x)$	$\nearrow$	162	$\searrow$	0	$\searrow$	-162	$\nearrow$

Since $f'(x)$ is zero only at the critical points, it is all positive or all negative in each interval between. For example, in the leftmost interval $(-\infty, 3)$, a sample value is $f'(-4) = 560 > 0$, so $f'(x)$ is positive in the whole interval, and we put $+$ in the first column next to $f'(x)$. The rest of the $f'(x)$ row is similar.

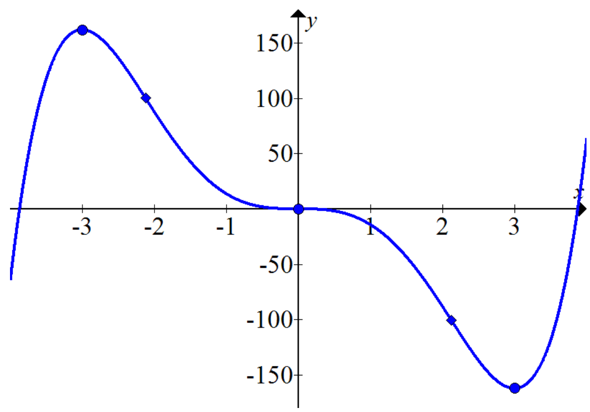
What does this mean for the graph $y = f(x)$? From §3.1, we know the critical points are candidates for local max/mins: hill tops or valley bottoms. Which is which? To the left of $x = -3$, we have $f'(x) > 0$ so $f(x)$ is increasing $\nearrow$; to the

Notes by Peter Magyar `magyar@math.msu.edu`

*Recall that $x \in (a,b)$ denotes the open interval, meaning $a < x < b$.

right, we have $f'(x) < 0$ so $f(x)$ is decreasing $\searrow$. Evidently, $x = 3$ is a local max, and the point $(-3, 162)$ is a hill top of the graph. Similarly, $(3, -162)$ is a valley.

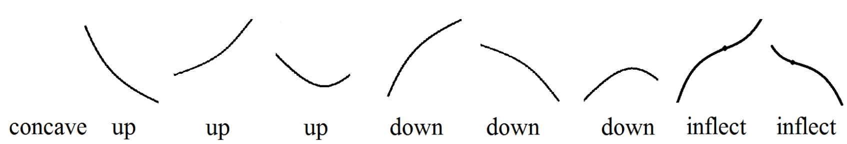
On the other hand, to the left *and* right of $x = 0$, we have $f'(x) < 0$, so $f(x)$ is decreasing on both sides: this means $x = 0$ is a stationary point where the graph levels out before continuing to descend. In fact, $f(x)$ is decreasing for $x \in [-3, 3]$, even though we only have $f'(x) \geq 0$, *not* $f'(x) > 0$ on this interval. We get a good picture of the graph:[†]



The reasoning in our example holds for any function:

> *First Derivative Test:* Let $f(x)$ be a function differentiable in a small interval around $x = c$, with $f'(c) = 0$.
>
> - If $f'(x) > 0$ for $x < c$ and $f'(x) < 0$ for $x > c$, then $x = c$ is a local maximum of $f(x)$.
> - If $f'(x) < 0$ for $x < c$ and $f'(x) > 0$ for $x > c$, then $x = c$ is a local minimum of $f(x)$.
> - If $f'(x)$ has the same sign on both sides of $x = c$, then $x = c$ is a stationary point of $f(x)$, not an extremal point.

Concavity. A more subtle feature of a graph is where it curves upward or downward. We say a graph is *concave up* near a point if it is part of a smiling curve $\smile$; and *concave down* if it is part of a frowning curve $\frown$. An *inflection point* is a special point where the graph wiggles, changing its concavity: a transition point between smiling and frowning $\sim$.[‡] Some examples:



In terms of the slope, concave up means that as x increases, the slope becomes less negative or more positive. For concave down, the slope becomes less positive or more negative.

[†] Also note that $f(x) = x^5 - 15x^3$ has only odd powers of x, so $f(-x) = -f(x)$. This means the graph has a 180° rotation symmetry, like a propeller. Such an $f(x)$ is called an *odd function.*

[‡] More generally, $y = f(x)$ is concave up if it lies below or on the secant line segment between any two points $(a, f(a))$ and $(b, f(b))$. For example, the absolute value graph $y = |x|$ has a "pointy smile" $\vee$ which lies below every secant line crossing the y-axis, and which contains every secant line on one side of the y-axis, so it is concave up.

Definition: Suppose the derivative $f'(x)$ is defined for x near c./

- $f(x)$ is *concave up* at $x = c$ if $f'(x)$ is *increasing* near $x = c$.
- $f(x)$ is *concave down* at $x = c$ if $f'(x)$ is *decreasing* near $x = c$.
- $f(x)$ has an *inflection point* at $x = c$ if $f'(x)$ has a local max or local min at $x = c$.

We can test for concavity using the second derivative $f''(x)$:

Concavity Theorem: Let $f(x)$ be a function.

- If $f''(x) > 0$ for all $x \in (a, b)$, then $f(x)$ is concave up over (a, b).
- If $f''(x) < 0$ for all $x \in (a, b)$, then $f(x)$ is concave down over (a, b).
- If $f''(c)=0$ and $f''(x)$ changes its sign at $x=c$, then $f(x)$ has an inflection point at $x = c$.

Proof. Applying the Increasing/Decreasing Theorem to the function $g(x) = f'(x)$, we get: if $g'(x) > 0$, then $g(x)$ is increasing. But $g'(x) = (f'(x))' = f''(x)$, so this means: if $f''(x) > 0$, then $f'(x)$ is increasing, and $f(x)$ is concave up. The proof of the second part is similar. The third part comes from applying the First Derivative Test to $g(x)$. Q.E.D.

EXAMPLE: Continuing the above $f(x) = x^5 - 15x^3$, $f'(x) = 5x^4 - 45x^2$, we have:

$$f''(x) \;=\; 20x^3 - 90x \;=\; 10x(2x^2{-}9).$$

The candidate inflection points are where $f''(x) = 0$, i.e $x = 0$ and $x = \pm\frac{3}{2}\sqrt{2} \approx \pm 2.12$, and the sign chart confirms the transitions in concavity at these points.

x		$-\frac{3}{2}\sqrt{2}$		0		$\frac{3}{2}\sqrt{2}$	
$f''(x)$	$-\,-$	0	$+\,+$	0	$-\,-$	0	$+\,+$
$f(x)$	$\frown$	$\frac{567}{8}\sqrt{2}$	$\smile$	0	$\frown$	$-\frac{567}{8}\sqrt{2}$	$\smile$

(I wrote double $-\,-$ and $+\,+$ just to make frowny and smiley faces: this is a good way to remember which is which.) This agrees with the features of our graph above, and it allows us to precisely determine the inflection points marked by small diamonds in the picture: $(-\frac{3}{2}\sqrt{2}, \frac{567}{8}\sqrt{2})$, $(\frac{3}{2}\sqrt{2}, -\frac{567}{8}\sqrt{2})$; and also $(0, 0)$, which is both a stationary critical point and an inflection point.

Critical Points and Concavity. There is one more use we can make of the second derivative. At a local max $x = c$, the slope changes from positive to negative, so the graph is concave down and $f''(c) < 0$; while at a local min it is concave up and $f''(x) > 0$. Thus, we can distinguish extremal points just from the sign of $f''(c)$.

Second Derivative Test: Let $f(x)$ be a function with $f''(x)$ continuous near $x = c$. Suppose $f'(c) = 0$.

- If $f''(c) < 0$, then $x = c$ is local maximum of $f(x)$.
- If $f''(c) > 0$, then $x = c$ is local minimum of $f(x)$.
- If $f''(x) = 0$, then this test fails, and $x = c$ might be a local max, a local min, or a stationary point.

Indeed, in our example, we have $f''(-3) = -270 < 0$ at the local max; $f''(0) = 0$ at the stationary point; and $f''(3) = 270 > 0$ at the local min.

Example. We will graph $f(x) = \dfrac{x^{2/3}}{(x-1)^2}$, going through the Method steps on the last page.

1. Using the Quotient and Chain Rules, and much simplification, we get:

$$f'(x) \;=\; \frac{\frac{2}{3}x^{-1/3}(1-x)^2 - x^{2/3}2(x-1)^1(x-1)'}{(1-x)^4} \;=\; -\frac{\frac{2}{3}(2x+1)}{(x-1)^3\,x^{1/3}}\;.$$

$$f''(x) \;=\; \frac{\frac{2}{3}(2)(1-x)^3x^{1/3} - \frac{2}{3}(2x+1)\left(3(x-1)^2x^{1/3} + (x-1)^3\frac{1}{3}x^{-2/3}\right)}{(1-x)^6\,x^{2/3}} \;=\; -\frac{\frac{2}{9}(14x^2+14x-1)}{(x-1)^4\,x^{4/3}}\;.$$

2. The two types of critical points are solutions of:
   - $f'(x) = 0$, when the numerator is zero: $\frac{2}{3}(2x+1) = 0$, i.e. $x = -\frac{1}{2}$
   - $f'(x) =$ undefined, when the denominator is zero, i.e. $x = 1$ and $x = 0$.

3. The sign chart looks like:

x		$-\frac{1}{2}$		0		1	
$f'(x)$	$+$	0	$-$	∞	$+$	∞	$-$
$f(x)$	$\nearrow$	$\frac{2}{9}\sqrt[3]{2}$	$\searrow$	0	$\nearrow$	∞	$\searrow$

   - $(-\frac{1}{2}, \frac{2}{9}\sqrt[3]{2}) \approx (-0.5, 0.28)$ is a local maximum (hill top dot in the pictue below). We could also see this by taking $f''(-\frac{1}{2}) = -\frac{32}{81}\sqrt[3]{2} < 0$.
   - $(0,0)$ is a local minimum, but instead of a flat valley bottom it is a sharp ravine (a cusp): instead of a horizontal tangent, the slope becomes infinte and the tangent line is vertical.
   - $x = 1$ is a vertical asymptote (dashed line in picture): since the denominator of $f(x)$ is zero, $f(1)$ is undefined and the function blows up to $\pm\infty$. Specifically, $f(x)$ is increasing to the left of $x = 1$, so the graph shoots up to $\lim_{x\to 1^-} f(x) = \infty$; and $f(x)$ is decreasing to the right of $x = 1$, so the graph shoots down from $\lim_{x\to 1^+} f(x) = \infty$.

4. The inflection points are solutions of $f''(x) = 0$, when the numerator is zero:

$$14x^2 + 14x - 1 = 0 \quad\Longleftrightarrow\quad x \;=\; \tfrac{-7\pm 3\sqrt{7}}{14} \;\approx\; -1.07,\; 0.07$$

   These are the small diamond points in the picture.

   Here the solutions of $f''(x) =$ undefined are just the vertical asymptote $x = 1$ of $f(x)$, and also the vertical asymptote $x = 0$ of $f'(x)$. Becasue $f'(a)$ is not defined, these are not considered inflection points, though the concavity does change.

5. The x and y-intercepts are both at $(0,0)$.

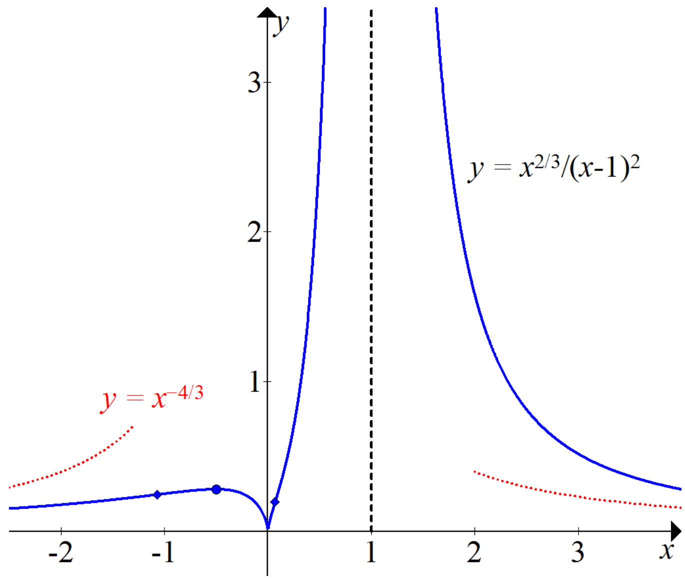
6. When x is a very large positive or negative number, x is almost the same as $x-1$ (compare 1000 and 999). We can approximate $f(x)$ by replacing $x-1$ with x:

$$f(x) \;=\; \frac{x^{2/3}}{(x-1)^2} \;\approx\; \frac{x^{2/3}}{x^2} \;=\; x^{-4/3} \;=\; \frac{1}{x\sqrt[3]{x}} \quad \text{for large } |x|.$$

   This simplified function is easy to graph (dotted curve in the picture), and the true graph $y = f(x)$ approaches this curve like an asymptote at the left and right ends of the x-axis.

7. This function does not have any symmetry.

8. Finally, the graph is:

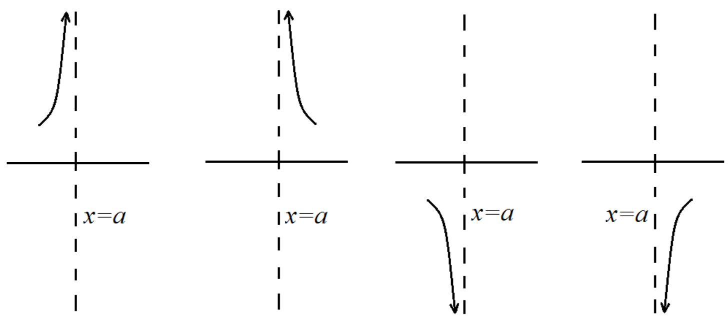


Method for Graphing

1. Determine the derivatives $f'(x)$ and $f''(x)$.
2. Solve $f'(x) = 0$ and $f'(x) = \text{undef}$ to find the critical points.
3. Sign table: $f'(x) > 0$ means $f(x)$ is $\nearrow$; $f'(x) < 0$ means $f(x)$ is $\searrow$. Classify critical points as: local max, local min, stationary, or vertical asymptote.
4. Solve $f''(x) = 0$ or undef for inflection pts. (Sign table usually not needed.)
5. Find the x-intercepts by solving $f(x) = 0$, and the y-intercept $(0, f(0))$.
6. Find the behavior as $x \to \pm\infty$ by taking the highest-order terms in $f(x)$
7. Check for symmetry: 180° rotation symmetry if $f(-x) = -f(x)$; or side-to-side reflection symmetry if $f(-x) = f(x)$.
8. Draw all the above features on the graph.

We will discuss Step 6 in §3.4. A very detailed Method chart is at the end of §3.5.

Vertical asymptotes. We say a curve has a line as an *asymptote* if, as the curve runs outward to infinity, it gets closer and closer to the line. "Closer and closer" reminds us of limits, and indeed we have seen that $x = a$ is a vertical asymptote of $y = f(x)$ whenever one of the following holds:



$$\lim_{x\to a^-} f(x) = \infty \quad \lim_{x\to a^+} f(x) = \infty \quad \lim_{x\to a^-} f(x) = -\infty \quad \lim_{x\to a^+} f(x) = -\infty.$$

As we saw in §1.5, ∞ has no meaning by itself; rather, the whole equation means that, as x gets closer to (but unequal to) a, the output $f(x)$ eventually becomes higher than any given bound B, such as $B = 100$ or 1000 or 1 billion. Similarly, a limit equals $-\infty$ when $f(x)$ becomes lower than $-B$ for any large B.

At the end of §3.3, we saw how a sign chart for $f'(x)$ can classify vertical asymptotes. We could do this with a sign chart for $f(x)$ itself, with no derivatives.

EXAMPLE: Let:

$$f(x) \;=\; \frac{x^2-6x+9}{x^3-6x^2+11x-6} \;=\; \frac{(x-3)^2}{(x-1)(x-2)(x-3)} \;=\; \frac{x-3}{(x-1)(x-2)}\,.$$

(To determine vertical asymptotes and intercepts, we always want $f(x)$ in factored* form.) In the original form, the denominator vanishes at $x = 3$, but we work with the cancelled form at right.

The function can only change its sign at points where $f(x) = 0$ (numerator = 0) or $f(x)$ is not defined (denominator = 0), that is, $x = 1, 2, 3$. In the interval $x \in (-\infty, 1)$, the sign is given by a sample point like $f(0) = \frac{-2}{(-1)(-3)} = -\frac{2}{3} < 0$, so $f(x)$ is negative; and similarly for the other intervals.

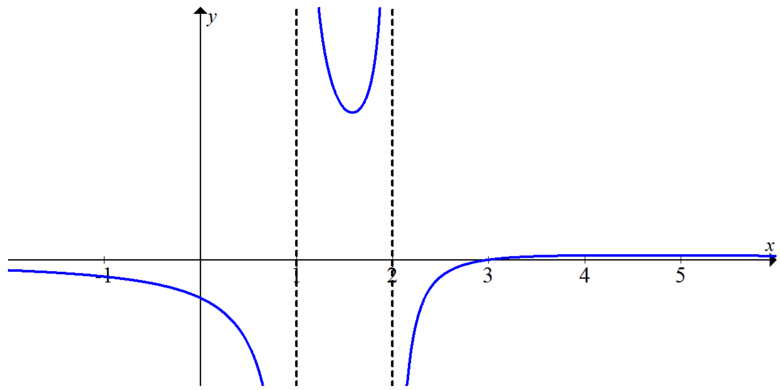
x		1		2		3	
$f(x)$	$-$	$\pm\infty$	$+$	$\pm\infty$	$-$	0	$+$

Each time x passes one of the sign-change candidates $x = a$, a factor $(x-a)$ changes from negative to positive, and $f(x)$ does indeed change sign.

Notes by Peter Magyar magyar@math.msu.edu

* To factor the bottom, we try linear factors $x - \frac{m}{n}$, where m is an integer factor of the constant coefficient 6, and n is an integer factor of the highest coefficient 1, so $n = \pm1, \pm2, \pm3, \pm6$ and $m = \pm1$. Trying $\frac{m}{n} = 1$, we find $x-1$ is a factor, since polynomial long division gives $x^3-6x^2+11x-6 = (x-1)(x^2-5x+6)$, and the quadratic is easy to factor. For a review of polynomial long division, see Khan Academy: www.khanacademy.org/math/algebra2/polynomial_and_rational/dividing_polynomials/v/polynomial-division.

Here $f(x) = \pm\infty$ just means the denominator vanishes and there is a vertical asymptote. The signs on each side of the asymptote show whether the graph shoots upward or downward: we have $\lim_{x\to 1^-} f(x) = -\infty$, $\lim_{x\to 1^+} f(x) = \infty$, $\lim_{x\to 2^-} f(x) = \infty$, $\lim_{x\to 2^+} f(x) = -\infty$.



Horzontal asymptotes. To understand the behavior of the graph over the left and right ends of the x-axis, we will need a new kind of limit in which x becomes larger and larger.

Definition:

- $\lim_{x\to\infty} f(x) = L$ means that $f(x)$ can be forced arbitrarily close to L, closer than any given $\varepsilon > 0$, by making $x > B$ for some B.
- $\lim_{x\to-\infty} f(x) = L$ means that $f(x)$ can be forced arbitrarily close to L, closer than any given $\varepsilon > 0$, by making $x < -B$ for some B.

Graphically, $\lim_{x\to\infty} f(x) = L$ means that toward the right of the x-axis, the graph $y = f(x)$ approaches the horizontal asymptote $y = L$; and similarly for $\lim_{x\to-\infty} f(x) = L$ toward the left. We can even have $\lim_{x\to\infty} f(x) = \infty$, which means that the graph goes off toward the upper right of the xy-plane in an unspecified way.

The most basic $x \to \infty$ limits are the power funcitons: for a positive real number power $p > 0$, we have:†

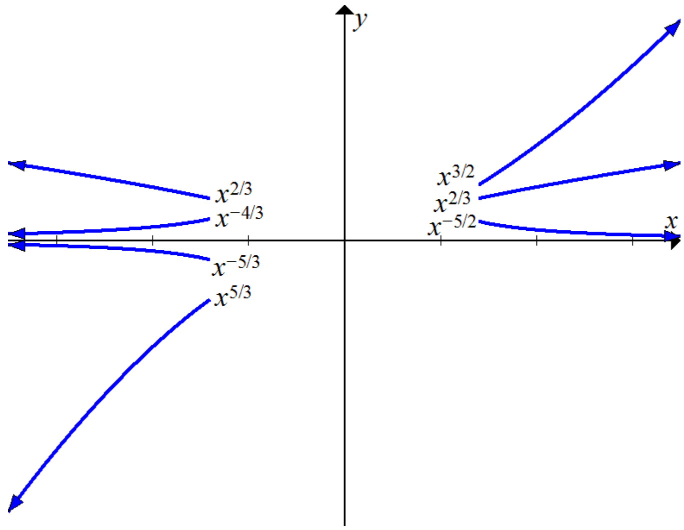
$$\lim_{x\to\infty} x^p = \infty, \qquad \lim_{x\to\infty} \frac{1}{x^p} = 0.$$

For $x \to -\infty$, consider the rational power $p = \frac{m}{n}$ where m, n are positive integers with n odd (perhaps $n = 1$); then:

$$\lim_{x\to-\infty} x^{m/n} = \begin{cases} \infty & \text{for } m \text{ even} \\ -\infty & \text{for } m \text{ odd,} \end{cases} \qquad \lim_{x\to-\infty} \frac{1}{x^{m/n}} = 0.$$

† *Proof:* For any large bound C, we can force $x^p > C$ if we take x so large that $x > C^{1/p}$. For any small error tolerance $\varepsilon > 0$, we can force $|\frac{1}{x^p} - 0| < \varepsilon$ if we take x so large that $x > (\frac{1}{\varepsilon})^{1/p}$.

For example:



Based on these, we can deduce the horizontal asymptotes for any rational function (quotient of polynomials).

EXAMPLE: Continuing $f(x) = \frac{x^2-6x+9}{x^3-6x^2+11x-6}$, does $y = f(x)$ have a horizontal asymptote? Informally, we can reason as follows. For large x (positive or negative), the value of x^2-6x+9 is relatively close to x^2: say for $x = 1000$, compare $x^2-6x+9 = 9{,}994{,}009$ and $x^2 = 1{,}000{,}000$. Thus we can approximate $x^2-6x+9 \approx x^2$, which we call the *highest term* of the polynomial. Also doing this for the denominator:

$$f(x) \;=\; \frac{x^2-6x+9}{x^3-6x^2+11x-6} \;\approx\; \frac{x^2}{x^3} \quad \text{for large } x.$$

Thus, $\lim\limits_{x\to\pm\infty} f(x) = \lim\limits_{x\to\pm\infty} \frac{x^2}{x^3} = \lim\limits_{x\to\pm\infty} \frac{1}{x} = 0$, and $y = f(x)$ has the horizontal asymptote $y = 0$ for $x \to \infty$ and $x \to -\infty$. In the graph we drew previously, the left and right ends do indeed approach the x-axis.

Formally, we can show this from the Limit Laws by dividing numerator and denominator by the highest term in the denominator:

$$\lim_{x\to\infty} f(x) \;=\; \lim_{x\to\infty} \frac{x^2-6x+9}{x^3-6x^2+11x-6} \;=\; \lim_{x\to\infty} \frac{x^2-6x+9}{x^3-6x^2+11x-6} \cdot \frac{\frac{1}{x^3}}{\frac{1}{x^3}}$$

$$= \lim_{x\to\infty} \frac{\frac{1}{x}-\frac{6}{x^2}+\frac{9}{x^3}}{1-\frac{6}{x}+\frac{11}{x^2}-\frac{6}{x^3}} \;=\; \frac{0-6(0)+9(0)}{1-6(0)+11(0)-6(0)} \;=\; 0.$$

Warning: The informal argument is the easiest way to understand these limits, but the formal argument (dividing by the highest term) might be required for full credit on a quiz or test.

EXAMPLE: For $f(x) = \frac{3x^2-x+9}{5x^2+2x-6}$, we take highest terms to get:

$$\lim_{x\to\infty} f(x) \;=\; \lim_{x\to\infty} \frac{3x^2-x+9}{5x^2+2x-6} \;=\; \lim_{x\to\infty} \frac{3x^2}{5x^2} \;=\; \frac{3}{5}.$$

Thus, $y = f(x)$ has horizontal asymptote $y = \frac{3}{5}$ toward the right. We similarly deduce $\lim\limits_{x\to-\infty} f(x) = \frac{3}{5}$, which means the same horizontal asymptote toward the left.

EXAMPLE: For

$$f(x) = \frac{x^2 + 3x^{7/2} - x^{-5}}{9x\sqrt{x} + 4x^2\sqrt{x}},$$

the terms in the denominator are $9xx^{1/2} = 9x^{3/2}$ and $4x^2x^{1/2} = 4x^{5/2}$, so the second is the highest term. Thus:

$$\lim_{x\to\infty} f(x) \;=\; \lim_{x\to\infty} \frac{x^2 + 3x^{7/2} - x^{-5}}{9x\sqrt{x} + 4x^2\sqrt{x}} \;=\; \lim_{x\to\infty} \frac{3x^{7/2}}{4x^{5/2}}$$

$$=\; \lim_{x\to\infty} \frac{3}{4}x^{7/2-5/2} \;=\; \lim_{x\to\infty} \frac{3}{4}x \;=\; \infty,$$

which means $y = f(x)$ has no horizontal asymptote. However, the approximation $f(x) \approx \frac{3}{4}x$ implies that the right end of the graph looks like a line with slope $\frac{3}{4}$. (See slant asymptotes in §3.5.) This function is not defined for $x < 0$, so there is no left end.

Man vs machine. In this section, we learn methods of drawing graphs by hand. The computer can do this much better simply by plotting many points, so why bother with our piddly sketches? One reason is that calculus tells us the critical areas of the graph to look at: the computer might default to showing us some uninteresting region which misses the key features. Another reason is to be able to check the answer for yourself.

This is part of a great danger for anyone who uses mathematics. If you let the computer do the thinking, not just the calculating, you are ready to blindly accept any bizarre wrong answer. Then one typo error can escalate until your scientific paper has to be retracted, your company's expenses are ten times what you predicted, your bridge collapses, your rocket crashes. Don't let it happen! Before you rely on the computer's answer, you must check it against reasonable expectations, qualitatively through a story or sketch, and quantitatively by plotting sample points.

Slant asymptote. This means a diagonal line $y = mx + b$ which is approached by a graph $y = f(x)$.* For example, consider the function:

$$f(x) \;=\; \frac{x^3 - 6x^2 + 11x - 6}{2x^2 - 8x}.$$

Recall from §3.3 that to find the large-scale behavior of $f(x)$ as $x \to \pm\infty$, we can approximate by the highest term in numerator and denominator: $f(x) \approx \frac{x^3}{2x^2} = \frac{1}{2}x$. Thus, the right and left ends of the graph look like lines with slope $\frac{1}{2}$.

However, the graph does not actually approach the line $y = \frac{1}{2}x$: there is a vertical shift, $y = \frac{1}{2}x + b$. To approximate better, and find the exact slant asymptote of $y = f(x)$, we perform polynomial long division:

$$\begin{array}{rl}
 & \tfrac{1}{2}x - 1 \;\text{ rem }\; 3x - 6 \\
2x^2 - 8x \,\big) & x^3 - 6x^2 + 11x - 6 \\
 & -\underline{(x^3 - 4x)} \\
 & \quad -2x^2 + 11x - 6 \\
 & -\underline{(-2x^3 + \;8x)} \\
 & \qquad 3x - 6
\end{array}$$

This means:

$$x^3 - 6x^2 + 11x - 6 \;=\; (\tfrac{1}{2}x - 1)(2x^2 - 8x) + (3x - 6),$$

so that:

$$f(x) \;=\; \frac{(\frac{1}{2}x - 1)(2x^2 - 8x) + (3x - 6)}{2x^2 - 8x} \;=\; \tfrac{1}{2}x - 1 + \frac{3x - 6}{2x^2 - 8x}.$$

That is, we have the approximation $f(x) \approx \frac{1}{2}x - 1$ with error term $\frac{3x-6}{2x^2-8x}$; but this term gets vanishingly small:

$$\lim_{x\to\pm\infty} \frac{3x - 6}{2x^2 - 8x} \;=\; \lim_{x\to\pm\infty} \frac{3x}{2x^2} \;=\; 0.$$

Notes by Peter Magyar `magyar@math.msu.edu`

*That is, the difference between them vanishes as x gets large: $\lim_{x\to\pm\infty} f(x) - (mx+b) = 0$.

That is, as x gets larger and larger, the error term gets smaller and smaller, and the graph $y = f(x)$ gets closer and closer to the line $y = \frac{1}{2}x - 1$. This is what we mean by a slant asymptote.

For a general rational function $f(x) = \frac{g(x)}{h(x)}$, a quotient of polynomials $g(x), h(x)$, we use polynomial long division to get $g(x) = q(x)h(x) + r(x)$ for a quotient polynomial $q(x)$ and a remainder polynomial $r(x)$ having lower powers of x than $h(x)$. Thus:

$$f(x) \;=\; \frac{g(x)}{h(x)} \;=\; \frac{q(x)h(x) + r(x)}{h(x)} \;=\; q(x) + \frac{r(x)}{h(x)}\,.$$

Since the numerator $r(x)$ is smaller than the denominator $h(x)$, we have $\lim_{x\to\pm\infty} \frac{r(x)}{h(x)} = 0$, and $y = f(x)$ gets closer and closer to the curve $y = q(x)$. If $q(x) = mx + b$, then $y = mx + b$ is a slant asymptote; otherwise, $y = q(x)$ is an *asymptotic curve* of $y = f(x)$.

Rational function example. Referring to the Method for Graphing at the end of this section, we apply the steps to the above function:

$$f(x) \;=\; \frac{x^3 - 6x^2 + 11x - 6}{2x^2 - 8x} \;=\; \frac{(x-1)(x-2)(x-3)}{2x(x-4)}\,.$$

1. We have:

$$f'(x) \;=\; \frac{x^4 - 8x^3 + 13x^2 + 12x - 24}{2x^2(x-4)^2}\,, \qquad f''(x) \;=\; -\frac{3(x^3 - 6x^2 + 24x - 32)}{x^3(x-4)^3}.$$

   The domain of $f(x)$ is all real numbers $x \neq 1, 4$, namely:

$$x \in (-\infty, 1) \cup (1, 4) \cup (4, \infty).$$

2. There is no neat way to solve $f'(x) = 0$. If we computer-plot the numerator $x^4 - 8x^3 + 13x^2 + 12x - 24$, we see 4 roots, which we can name $x = a_1, \ldots, a_4$, approximately at:

$$a_1 \approx -1.26, \quad a_2 \approx 1.39, \quad a_3 \approx 2.61, \quad a_4 \approx 5.26\,.$$

   In §3.8, we will learn Newton's Method to zero in on such approximate solutions when algebraic ones are not available.

   The other critical points are solutions of $f'(x) = $ undefined, namely the roots of the denominator $x = 0$ and 4.

3. The sign chart is:

x		a_1 -1.26		0		a_2 1.39		a_3 2.61		4		a_4 5.26	
$f'(x)$	+	0	−	$\pm\infty$	−	0	+	0	−	$\pm\infty$	−	0	+
$f(x)$	↗	−2.37	↘	$\pm\infty$	↘	−0.05	↗	0.05	↘	$\pm\infty$	↘	2.37	↗
		max		asym		min		max		asym		min	

4. To solve $f''(x) = 0$, a computer-plot of the numerator $x^3 - 6x^2 + 24x - 32$ appears to show a single root at $x = 2$. To check this, we do polynomial long division $(x^3 - 6x^2 + 24x - 32) \div (x-2)$ to show the numerator is $(x-2)(x^2 - 4x + 16)$,

where the quadratic factor has no real number roots. Thus $x = 2$ is the only inflection point. (Solving $f''(x) = \text{undef}$ just gives the vertical asymptotes.)

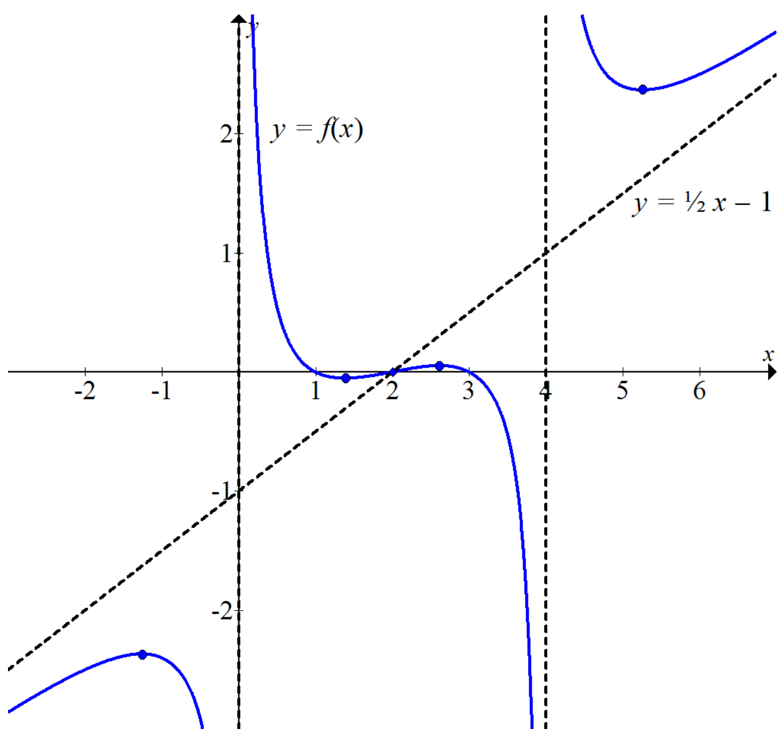
We do not need a sign chart for $f''(x)$, since the concavity seen in the picture below is forced by the known critical and inflection points: anything else would lead to more wiggles.

5. Solving $f(x) = 0$ gives the x-intercepts $x = 1, 2, 3$. There is no y-intercept, since the y-axis is a vertical asymptote.

6. The slant asymptote is $y = \frac{1}{2}x - 1$, computed at the beginning of this section.

7. This function does not have any of the standard symmetries in the Method. However, the graph reveals a 180° rotation symmetry around the point $(2, 0)$. This is equivalent to the equation $f(4-x) = -f(x)$, which can be shown from the factored form.

8. The graph is:



Trigonometric example. Apply the Method at the end to: $s(x) = x - 2\sin(x)$.

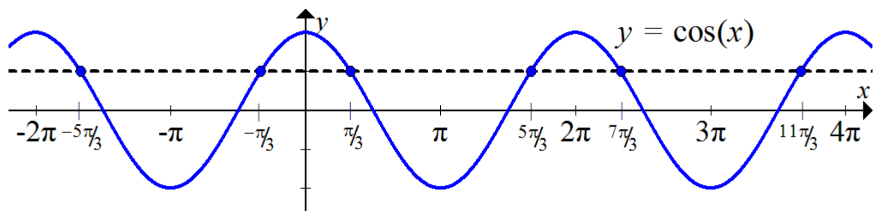
1. We have: $s'(x) = 1 - 2\cos(x)$ and $s''(x) = 2\sin(x)$. (See §2.4.)

   The domain is all real numbers: $x \in (\infty, \infty)$.

2. The critical points are solutions of $s'(x) = 1 - 2\cos(x) = 0$, so $\cos(x) = \frac{1}{2}$, and $x = 60° = \frac{1}{3}\pi$, or $x = 2\pi - \frac{1}{3}\pi = \frac{5}{3}\pi$, or any shift of these by a multiple of 2π:

   $$x = \tfrac{1}{3}\pi \pm 2n\pi \quad \text{and} \quad \tfrac{5}{3}\pi \pm 2n\pi \quad \text{for } n = 0, 1, 2 \ldots.$$

   You can see this on the graph of $\cos(x)$:

   There are no points with $s'(x) =$ undefined.

3. The sign chart for $s'(x)$ is periodic (repeating):

x	$\cdots$		$-\frac{5}{3}\pi$		$-\frac{1}{3}\pi$		$\frac{1}{3}\pi$		$\frac{5}{3}\pi$		$\frac{7}{3}\pi$		$\frac{11}{3}\pi$		$\cdots$
$s'(x)$	$\cdots$	$-$	0	$+$	0	$-$	0	$+$	0	$-$	0	$+$	0	$-$	$\cdots$
$s(x)$	$\cdots$	$\searrow$	$-\frac{5}{3}\pi-\sqrt{3}$	$\nearrow$	$-\frac{1}{3}\pi+\sqrt{3}$	$\searrow$	$\frac{1}{3}\pi-\sqrt{3}$	$\nearrow$	$\frac{5}{3}\pi+\sqrt{3}$	$\searrow$	$\frac{7}{3}\pi-\sqrt{3}$	$\nearrow$	$\frac{11}{3}\pi+\sqrt{3}$	$\searrow$	$\cdots$
	$\cdots$		min		max		min		max		min		max		$\cdots$

4. The inflection points are solutions of $s''(x) = 2\sin(x) = 0$, or $x = n\pi$ for any integer n. Every multiple of π is an inflection point of $y = s(x)$.

5. The point $(0, s(0)) = (0, 0)$ is an x and y-intercept. From the graph, we can see that there are two more x-intercepts, but we have no way to find them exactly. (We can approximate by Newton's Method §3.8.)

6. The large-scale behavior can be approximated by taking the highest or largest term: $s(x) \approx x$. However, the line $y = x$ is not a slant asymptote, because $s(x)$ oscillates above and below this line, without getting closer and closer.

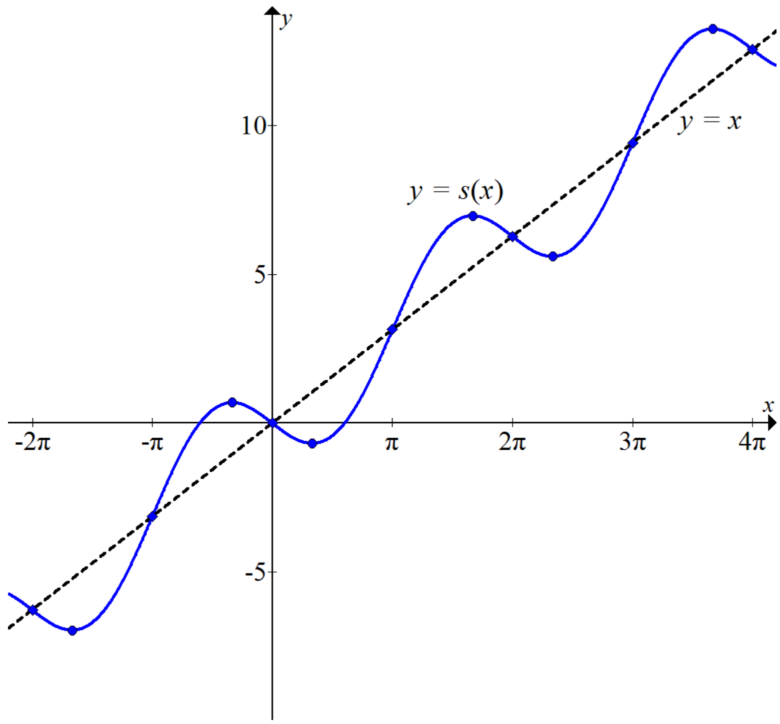
7. This is an odd function, since:

   $$s(-x) = (-x) - 2\sin(-x) = -x + 2\sin(x) = -s(x).$$

   Thus, the graph has 180° rotation symmetry around the origin.

   This function is not periodic, since $s(x+2\pi) \neq s(x)$, so the graph does not have the shift-sideways translation symmetry. However, we do have $s(x+2\pi) = s(x)+2\pi$, so the graph can be moved to itself by shifting sideways and up!

8. The graph is:



Method for Graphing (detailed)

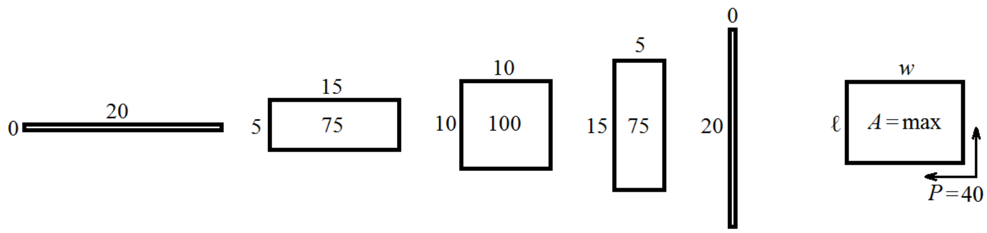
1. Determine the derivatives $f'(x)$ and $f''(x)$ with Derivative Rules.
   Determine the domain of $f(x)$: for what x the formula makes sense.
2. Solve $f'(x) = 0$ and $f'(x) =$ undef to find the critical points.
3. Make a sign table for $f'(x)$ to classify each critical point $x = a$:

		$x<a$	$x=a$	$x>a$
local max $\frown$	$f'(x)$	$+$	0	$-$
	$f(x)$	$\nearrow$	$f(a)$	$\searrow$
local min $\smile$	$f'(x)$	$-$	0	$+$
	$f(x)$	$\searrow$	$f(a)$	$\nearrow$
local max $\wedge$	$f'(x)$	$+$	undef	$-$
	$f(x)$	$\nearrow$	$f(a)$	$\searrow$
local min $\vee$	$f'(x)$	$-$	undef	$+$
	$f(x)$	$\searrow$	$f(a)$	$\nearrow$
vert asymp $\nearrow\vert\nwarrow$	$f'(x)$	$+$	$\frac{1}{0}$	$-$
	$f(x)$	$+$	$\frac{1}{0}$	$+$
vert asymp $\nearrow\vert\swarrow$	$f'(x)$	$+$	$\frac{1}{0}$	$+$
	$f(x)$	$+$	$\frac{1}{0}$	$-$
vert asymp $\searrow\vert\nwarrow$	$f'(x)$	$-$	$\frac{1}{0}$	$-$
	$f(x)$	$-$	$\frac{1}{0}$	$+$
vert asymp $\searrow\vert\swarrow$	$f'(x)$	$-$	$\frac{1}{0}$	$+$
	$f(x)$	$-$	$\frac{1}{0}$	$-$

   Here $f(a)$ means the output value is defined; and $\frac{1}{0}$ means a zero denominator at $x = a$ produces $\pm\infty$ values. There other possibilities if $x = a$ is a discontinuity (see §1.8).

4. Solve $f''(x) = 0$ or undef to find inflection points $x = a$; we also require that $f'(a)$ exists and is a local max/min of $f'(x)$. Make a sign table for $f''(x)$ if concavity is needed: $f''(x) > 0$ means concave up (smiling), $f''(x) < 0$ means concave down (frowning).
5. Solve $f(x) = 0$ to find the x-intercepts; and compute the y-intercept $(0, f(0))$.
6. Find the behavior as $x \to \pm\infty$.
   - Approximate by highest terms on top and bottom to get $f(x) \approx cx^p$.
   - For a better approximation of a rational function $f(x) = \frac{g(x)}{h(x)}$, use polynomial long division to get $f(x) = q(x) + \frac{r(x)}{h(x)}$.
     If $f(x) = mx + b + \frac{r(x)}{h(x)}$, then $y = mx + b$ is a slant asymptote.
     In general, $y = f(x)$ asymtotically approaches $y = q(x)$ as $x \to \pm\infty$.
7. Check for symmetries: ways to move the graph onto itself.
   - Side-to-side reflection symmetry for even function $f(-x) = f(x)$.
     EXAMPLES: $x^2{+}3$, x^4, $\cos(x)$
   - 180° rotation symmetry for odd function $f(-x) = -f(x)$.
     EXAMPLES: $2x$, x^3, $\sin(x)$
   - Shift-sideways translation symmetry for periodic $f(x{+}c) = f(x)$.
     EXAMPLES: $\cos(x{+}2\pi) = \cos(x)$, $\tan(x{+}\pi) = \tan(x)$.
8. Draw all the above features on the graph.

Rectangle example. Suppose we have 40 meters of fence to make a rectangular corral. What length and width will fence off the largest area? The range of possiblilities is illustrated below:



It appears that the square with length and width $\ell = w = 10$ gives the maximum area $A = \ell w = 100$ m^2. To prove this algebraically, we note that the perimeter is constant, $P = 2\ell + 2w = 40$; so the length controls the width and also the area:

$$w = \tfrac{1}{2}(40 - 2\ell) = 20 - \ell, \qquad A = \ell w = \ell(20 - \ell) = 20\ell - \ell^2\,.$$

That is, the quantity we aim to maximize, A, is a function of the variable ℓ, which is allowed to vary between $\ell = 0$ and $\ell = 20$ (corresponding to $w = 0$). This is a familiar problem: find the absolute maximum of

$$A(\ell) = 20\ell - \ell^2 \quad \text{over the interval } \ell \in [0, 20].$$

The critical points are given by $\frac{dA}{d\ell} = 20 - 2\ell = 0$, i.e. $\ell = 10$ with output $A(10) = 100$, and the endpoint outputs are $A(0) = 0$, $A(20) = 0$. The largest of these is the absolute maximum: $\ell = 10$ with $A(\ell) = 100$; also $w = 20 - \ell = 10$.

Method for optimization. We aim to find the maximum or minimum possible value of a target quantity within the constraints of a (usually geometric) situation.

1. Draw a picture labeled with numerical constant values and with letters for varying quantities, including: *controlling* variables to determine the shape; *constrained* variables required to have a fixed value; the *target* variable we aim to maximize or minimize.

2. Write equations relating variables according to the geometry of the picture.

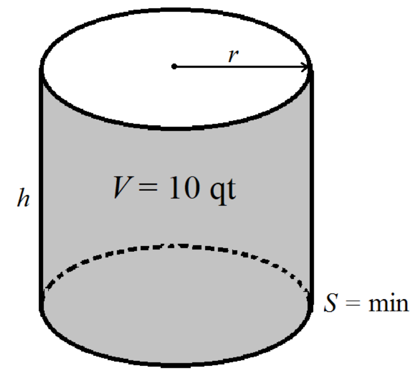
3. Choose one of the controlling variables (say, x) as the *independent* variable, and write all other variables as functions of it by solving the above equations. Also determine the relevant domain $x \in [a, b]$, which is usually restricted by requiring all lengths to be positive.

4. Find the absolute maximum/minimum of the target variable over its domain, say $T = T(x)$ over $x \in [a, b]$. That is, solve $T'(x) = 0$ or undef, to find the critical points $x = c_1, c_2, \ldots,$ as well as the endpoints $x = a, b$. Take the output values $T(x)$ at these candidate points: the largest/smallest output is the desired maximum/minimum.

5. If needed, find values of the other variables at the optimum x. Make sure the answer is physically plausible to check for mistakes.

Notes by Peter Magyar `magyar@math.msu.edu`

Bucket example. Consider a 10-quart bucket with cylindrical sides and circular bottom. What radius and height will minimize surface area of sides and bottom?

1. 

   The *target* variable is the surface area S (square inches), to be minimized. The *controlling* variables are radius r (inches) and height h (inches). The constant volume is expressed by the *constrained* variable $V = 10$ quarts; to make this comparable to the other variables, we must convert to $V = 577.5$ cubic inches.

2. Equations. The volume V is the base area πr^2 times the height h. For the surface S: the sides, if unrolled, form a rectangle with the same height h as the cylinder, and width equal to the perimeter of the bottom, $2\pi r$; and we also add the bottom area πr^2. Thus:

$$V \;=\; \pi r^2 h = 577.5\,, \qquad S \;=\; \pi r^2 + 2\pi r h \;=\; \text{min}.$$

3. Do we choose r or h as the *indendent* variable? Here r is harder to solve for, so we make it independent and solve for the other variables instead:

$$h \;=\; \frac{577.5}{\pi r^2}, \qquad S \;=\; \pi r^2 + 2\pi r\,\frac{577.5}{\pi r^2} \;=\; \pi r^2 + \frac{1155}{r}.$$

   The only restriction on r is $r > 0$. (Radius can be huge if height is correspondingly tiny: this is clearly not optimal, but still possible.) Thus, the domain is the open interval $r \in (0, \infty)$.

4. We must find the absolute minimum of $S(r) = \pi r^2 + \frac{1155}{r}$ over $r \in (0, \infty)$. To find the critical points:

$$\frac{dS}{dr} \;=\; 2\pi r - \frac{1155}{r^2} \;=\; 0 \quad\Longrightarrow\quad 2\pi r = \frac{1155}{r^2} \quad\Longrightarrow\quad r \;=\; \sqrt[3]{\frac{1155}{2\pi}} \;\approx 5.68\,.$$

   This is the only critical point, with output value $S(r) \approx 304\,$.

   Since the endpoint values $S(0)$ and $S(\infty)$ are not defined, we must consider the limiting values near these points: $\lim_{r\to 0^+} S(r) = \infty$ and $\lim_{r\to\infty} S(r) = \infty$. This means $S(r)$ has no absolute maximum, but can get as large as desired if we make r large or small enough.

   The remaining candidate must be the absolute minimum point, $r = \sqrt[3]{\frac{1155}{2\pi}}$.

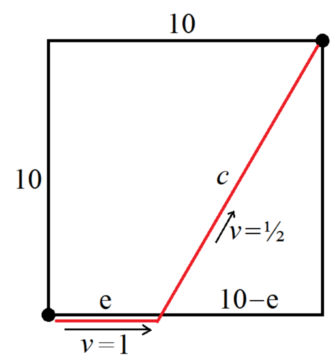
5. At the minimum point, the other variable is:

$$h \;=\; \frac{577.5}{\pi\left(\sqrt[3]{\frac{1155}{2\pi}}\right)^2} \;\approx\; 5.68\,.$$

In fact, we can simplify to show $h = r$, meaning the optimal bucket is twice as wide as it is high.

This is plausible. However, an actual 10-quart bucket has dimensions about as wide as it is high, about $r = \frac{1}{2}h = 4.5$ in, which uses more plastic than necessary. Try to explain what other factors might influence the design.

Ants example. A line of ants marches across a 10cm $\times$ 10cm square of carpet from the lower left to the upper right corner (where someone dropped a jellybean). Part of their path is along the edge next to the carpet, where their speed is 1 cm/sec, and part diagonally across the carpet, where their speed is $\frac{1}{2}$ cm/sec. What path should they take along the edge before entering the carpet, so as to minimize (a) the total distance; and (b) the total travel time.

1. 

Controlling variables are e, the distance traveled along the edge, and c, the distance traveled across the carpet. The target variable to minimize for each question is: (a) total distance D in cm; and (b) total time T in sec.

2. Equations:

$$c^2 = 10^2 + (10{-}e)^2\,, \qquad D = e + c\,.$$

Also, we know speed $\times$ time = distance, so time = distance/speed. The travel time along the edge is $e/1 = e$, along the carpet $c/\frac{1}{2} = 2c$, with total:

$$T \;=\; e + 2c\,.$$

3. The obvious independent variable is e, since we can easily write the other variables in terms of it, including the target variables:

$$c \;=\; \sqrt{10^2{+}(10{-}e)^2} \;=\; \sqrt{200{-}20e{+}e^2}\,,$$

$$D \;=\; e + \sqrt{200{-}20e{+}e^2}\,, \qquad T \;=\; e + 2\sqrt{200{-}20e{+}e^2}\,.$$

The relevant domain is $e \in [0, 10]$.

4. For question (a), the critical points are given by:

$$\frac{dD}{de} = 1+\tfrac{1}{2}(200{-}20e{+}e^2)^{-1/2}(200{-}20e{+}e^2)' = 1-\frac{10-e}{\sqrt{200{-}20e{+}e^2}} = 0\,,^*$$

which reduces to $\sqrt{200{-}20e{+}e^2} = 10-e$, then to $200{-}20e{+}e^2 = (10{-}e)^2$, which cancels to the impossible equation $200 = 100$. Thus, there are *no* critical points, and the absolute minimum must be one of the endpoints. Since $D(0) = 10\sqrt{2} \approx 14.4 < D(10) = 20$, the minimum is at $e = 0$.

For question (b), the critical points are given by:

$$\frac{dT}{de} = 1-\frac{2(10{-}e)}{\sqrt{200{-}20e{+}e^2}} = 0 \implies \sqrt{200{-}20e{+}e^2} = 20-2e$$

$$\implies 200-20e+e^2 = (20-2e)^2 \implies 3e^2-60e+200 = 0\,.$$

The Quadratic Formula then gives:

$$e = \tfrac{60\pm\sqrt{60^2-4(3)(200)}}{2(3)} = 10 \pm \tfrac{10}{3}\sqrt{3} \approx 4.2,\ 15.8\,.$$

The second solution is outside the domain $e \in [0,10]$, so the only relevant critical point is $e = 10-\frac{10}{3}\sqrt{3} \approx 4.2$, with value $T(e) = 10+10\sqrt{3} \approx 27.3\,.$ Comparing to endpoints $T(0) = 20\sqrt{2} \approx 28.3$ and $T(10) = 30$, we find the absolute minimum at $e = 10{-}\frac{10}{3}\sqrt{3} \approx 4.2$ with $T(e) = 10{+}10\sqrt{3} \approx 27.3$.

5. For question (a), the minimum distance at $e = 0$ is obvious in retrospect: the straight diagonal is the shortest path between opposite corners.

   For question (b), the minimum time is about $T(4.2) \approx 27.3$ sec: that is, at a speed between 0.5 and 1 cm/sec, the ants can cross the 10 cm $\times$ 10 cm square in about 27 sec, which is reasonable. This is a slight saving over the straight diagonal path, which takes about 28 sec. (This assumes they move at carpet speed along the right edge of the square; if they moved at floor speed, they would do much better to go around the carpet, at 20 sec.)

A line of ants will usually find the minimum distance path over a landscape by gradually tightening their curves; what do you think they would do in this case?

Maximizing profit. The Acme Company produces widgets for \$10 each and sells them for s dollars each. The number of widgets sold is modeled by the market demand function $m(s) = 100-s$: for example, if they charge \$25, customers will buy $m(25) = 75$ widgets, but price \$100 is too high for the market: $m(100) = 0$.

PROBLEM: What selling price s will maximize total profits?

The independent variable is $s \in [10,100]$. Profit per widget is $s-10$. Total profit is $P(s) = m(s)\,(s{-}10) = (100{-}s)(s{-}10) = -1000+110s-s^2$. The critical point $P'(s) = 110-2s = 0$ is $s = 55$, which is clearly the maximum point, since $P(55) = 2025$ but the endpoints produce $P(10) = P(100) = 0$. Thus the most profitable selling price is $s = \$55$.

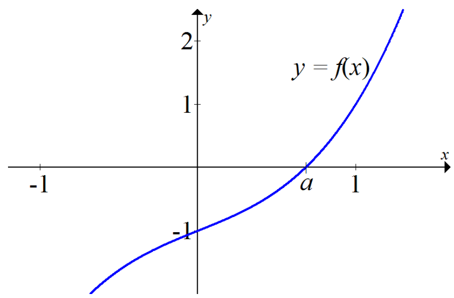
* Note $\frac{dD}{de}$ is defined over the whole domain $e \in [0,10]$, since $200{-}20e{+}e^2 = 10^2{+}(10{-}e)^2 > 0$.

Roots of equations. We frequently need to solve equations for which there is no neat algebraic solution, such as:

$$f(x) \ = \ x^3 + x - 1 \ = \ 0\,.$$

In this case, the best we can ask is an approximate solution, accurate to a specified number of decimal places, and this is all we need for any practical purpose.

We can start with a computer graph of $y = f(x)$, which is just a display of many plotted points $(x, f(x))$:



A solution of $f(x) = 0$ is an x-intercept of the graph, and we see one,* call it $x = a$, close to $x = 0.5$; that is, our first estimate is $a \approx 0.5$. Computing:

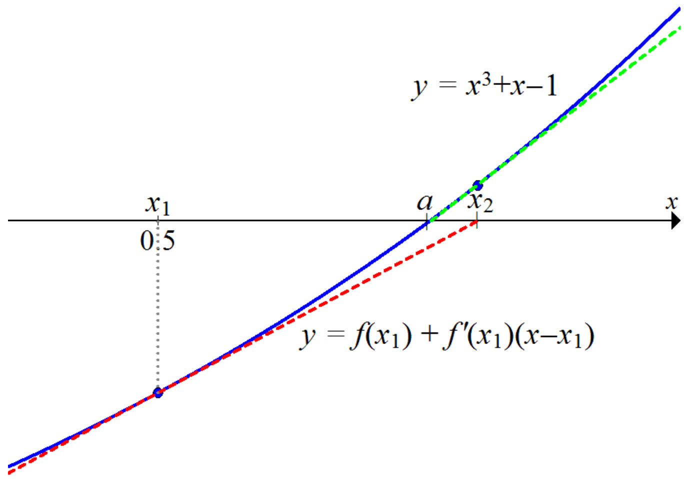
$$f(0.5) = -0.375 < 0, \quad f(0.6) = -0.184 < 0, \quad f(0.7) = 0.043 > 0,$$

the Intermediate Value Theorem (§1.8) guarantees a solution $0.6 < a < 0.7$; thus we can improve our estimate to $a \approx 0.6$. We could add a decimal place by checking $f(0.61), f(0.62), \ldots, f(0.69)$ to see where the values change from negative to positive, but this is clearly very tedious and inefficient.

Newton's Method is an amazingly efficient way to refine an approximate solution to get more and more accurate ones, until the required accuracy is reached. Let us call our first estimate $x_1 = 0.5$. We are seeking the true solution $x = a$, the x-intercept of $y = f(x)$. As in §2.9, let us approximate $y = f(x)$ by its tangent line at our initial point at $(x_1, f(x_1))$, namely $y = f(x_1) + f'(x_1)(x{-}x_1)$:

Notes by Peter Magyar `magyar@math.msu.edu`

* How do we know there is no other solution $x = b$? If there were, Rolle's Theorem (§3.2) says that there would be some $x = c \in (a, b)$ with $f'(c) = 0$, namely a hill or valley of $y = f(x)$. But $f'(x) = 3x^2 + 1 = 0$ clearly has no solutions, so $y = f(x)$ has no hills or valleys, and there cannot exist another solution $x = b$.



You can see how the tangent line (in red) is very close to the graph near $x = x_1$, and fairly close even near the true solution $x = a$. We cannot solve for the x-intercept of $y = f(x)$, but we can find the x-intercept of the line, denoted $x = x_2$:

$$f(x_1) + f'(x_1)(x-x_1) \;=\; 0 \qquad \Longrightarrow \qquad x \;=\; x_2 \;=\; x_1 - \frac{f(x_1)}{f'(x_1)}\,.$$

This solution x_2 is not exactly a, but it is closer than the initial estimate x_1.

Now we can iterate (green line), repeating the same computation starting with x_2 instead of x_1. The result is:

$$x_3 \;=\; x_2 - \frac{f(x_2)}{f'(x_2)}\,,$$

which is much closer to a; then $x_4 = x_3 - \frac{f(x_3)}{f'(x_3)}$; and repeating the same way we get the following spreadsheet, computing to 3 decimal places:

n	x_n	$f(x_n)$	$f'(x_n)$	x_{n+1}
1	0.500	−0.375	1.750	0.714
2	0.714	0.079	2.531	0.683
3	0.683	0.002	2.400	0.682
4	0.682	0.000	2.397	0.682
5	0.682	0.000	2.397	0.682

The x_n's will continue as real numbers to converge closer and closer to a, but since we do not see any difference in our 3 decimal places after x_4, there is no point in continuing. We already have our answer within the specified accuracy:

$$a \approx 0.682 \quad \text{accurate to 3 decimal places.}$$

In the table, $f(x_4) \approx 0.000$ is indeed an approximate solution to $f(x) = 0$.

Newton's Method. We wish to solve an equation $f(x) = 0$, with the true solution $x = a$ fairly close to an intial estimate $a \approx x_1$, and the final approximation $a \approx x_n$ accurate to a specified number of decimal places.

1. Using a calcuator, spreadsheet, or computer algebra system, compute $x_2, x_3, \ldots$ according to the formula:

$$x_{n+1} \;=\; x_n - \frac{f(x_n)}{f'(x_n)},$$

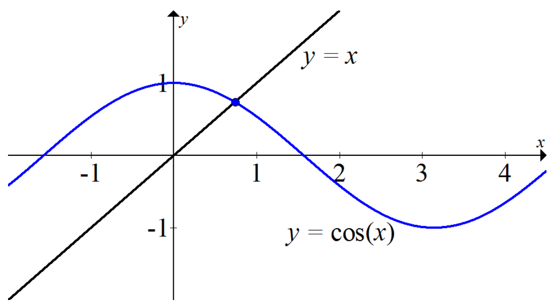
computing with *at least* the specified accuracy (number of decimal places).

2. Stop once $x_n \approx x_{n+1}$ are the same up to the given accuracy. The final approximation is $a \approx x_n$.

Trigonometric equation: Let us solve, to 3 decimal places, the equation:

$$\cos(x) = x\,.$$

(As always in Calculus, we assume x is in radians: see §2.5 end.)



Looking at the graph, we see that there is a unique solution somewhere around $x_1 = 1$. This seems different from the previous case, since we seek the intersection of two graphs rather than the x-intercept of a single graph; but we can simply rewrite the equation as $f(x) = x - \cos(x) = 0$. Newton's Method gives:

$$x_{n+1} \;=\; x_n - \frac{x_n - \cos(x_n)}{1+\sin(x_n)}\,,$$

x_1	x_2	x_3	x_4
1.000	0.750	0.739	0.739

That is, the solution is $a \approx 0.739$ to 3 places.

Numerical roots. The number $\sqrt{2}$ is a "known value": a calculator can immediately tell us that $\sqrt{2} = 1.41421356\ldots$. But just how does the calculator know this? Newton's Method, that's how!

By definition, $\sqrt{2}$ is the solution of $x^2 = 2$, or $f(x) = x^2 - 2 = 0$. Starting with $x_1 = 1$, the Method gives $x_{n+1} = x_n - \frac{x_n^2-2}{2x_n}$ and:

x_1	1.00000000
x_2	1.50000000
x_3	1.41666667
x_4	1.41421569
x_5	1.41421356
x_6	1.41421356

Here we see the power of the Method: with just a couple of dozen $+, -, \times, \div$ calculator operations, it converged from 0 places to 8 places of accuracy.

We could also do the Method with fractions rather than decimals to get very accurate fractional approximations of $\sqrt{2}$:

x_1	1
x_2	3/2
x_3	17/12
x_3	577/408

Already $x_3 = \frac{17}{12}$ is a very good approximation, since $(\frac{17}{12})^2 = \frac{289}{144} = 2\frac{1}{144}$, very close to 2. However, no fraction or finite decimal can give $\sqrt{2}$ exactly: it is known to be an *irrational* number.

Antiderivatives

Reversing differentiation. In many problems, especially physical ones, we are interested in some function $F(x)$, but we only know its derivative $F'(x) = f(x)$. We need to reverse the differentiation process to find the original $F(x)$, the *antiderivative* of $f(x)$, also called the *primitive* of $f(x)$.

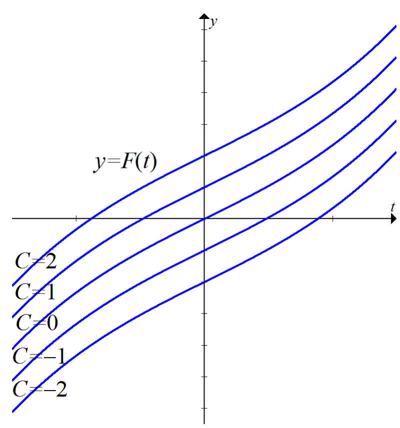
For example, suppose $F(t)$ represents the height of an object at time t, but we only know the velocity:

$$F'(t) = f(t) = t^2 + 2.$$

What was the original $F(t)$? Recalling our Basic Derivative $(t^n)' = nt^{n-1}$, we realize that $(\frac{1}{3}t^3)' = \frac{1}{3}(3t^2) = t^2$, so $F(t) = \frac{1}{3}t^3 + 2t$ works. But this is not the only possible answer because any constant term C disappears in the derivative, so the general answer is:

$$F(t) \;=\; \tfrac{1}{3}t^3 + 2t + C.$$

This family of functions is called the *general antiderivative*:



The non-uniqueness of $F(t)$ means that the velocity alone does not determine the height. But if we know the height at just one time, for example the initial height $F(0) = 5$, then we can adjust the constant C in a unique way to satisfy this requirement:

$$F(0) \;=\; \tfrac{1}{3}(0^3) + 2(0) + C \;=\; 5 \quad\Longrightarrow\quad C = 5.$$

That is, $F(t) = \frac{1}{3}t^3 + 2t + 5$ is the unique function with $F'(t) = t^2 + 2$ and $F(0) = 5$. We have solved an *initial value problem.* (See the Ballistic Equation, end of §2.7.)

Generalizing we have:

> *Definition.* $F(x)$ is an *antiderivative* of $f(x)$ means $F'(x) = f(x)$.
>
> *Antiderivative Theorem.* Assume $F(x)$ is some particular antiderivative of $f(x)$ for $x \in (a, b)$. Then:
>
> (a) The general antiderivative of $f(x)$ is $F(x) + C$ for any constant C. There are no other antiderivatives of $f(x)$.
>
> (b) For any $c \in (a, b)$ and any A, there is a unique antiderivative $\tilde{F}(x)$ satisfying the intial value problem $\tilde{F}'(x) = f(x)$ and $\tilde{F}(c) = A$.

Notes by Peter Magyar magyar@math.msu.edu

Proof. (a) This just rephrases §3.2 Uniqueness Theorem (b). That is, if $\tilde{F}(x)$ is any antiderivative, i.e. any function with $\tilde{F}'(x) = F'(x) = f(x)$, then the Uniqueness Theorem guarantees $\tilde{F}(x) = F(x) + C$ for some constant C.

(b) We are given a specific antiderivative $F(x)$ by hypothesis, so by part (a), a general antiderivative is $\tilde{F}(x) = F(x) + C$. If we require $\tilde{F}(c) = F(c) + C = A$, this determines C uniquely as $C = A - F(c)$, so we can only have $\tilde{F}(x) = F(x) + A - F(c)$. (See also §3.2 Uniqueness Theorem (c).) Q.E.D.

Antidifferentiation. This means the process of finding antiderivatives by reversing the rules for derivatives from §2.3–2.4. For every Basic Derivative of the form $F'(x) = f(x)$, we have a reverse Basic Antiderivative:

$F(x)$	$F'(x) = f(x)$
x^n	nx^{n-1}
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$\tan(x)$	$\sec^2(x)$
$\sec(x)$	$\tan(x)\sec(x)$

$\Longrightarrow$

$f(x)$	$F(x)$
x^n	$\frac{1}{n+1}x^{n+1} + C$
$\cos(x)$	$\sin(x) + C$
$\sin(x)$	$-\cos(x) + C$
$\sec^2(x)$	$\tan(x) + C$
$\tan(x)\sec(x)$	$\sec(x) + C$

Each general antiderivative has an arbitrary constant term C. Notice we do *not* know an antiderivative for $f(x) = \frac{1}{x} = x^{-1}$, since the formula $\frac{1}{-1+1}x^0$ does not make sense.

We can also reverse the Derivative Rules. Since the derivative of a sum is the sum of derivatives, the same is true for antiderivatives, and similarly for differences and constant multiples:

$$\begin{aligned} f(x) &= 7x^3 - x\sqrt{x} + \frac{3}{x^2} - \frac{4}{\cos^2(x)} \\ &= 7x^3 - x^{3/2} + 3x^{-2} - 4\sec^2(x), \end{aligned}$$

$$\begin{aligned} F(x) &= 7(\tfrac{1}{4}x^4) - \tfrac{1}{5/2}x^{5/2} + 3(\tfrac{1}{-1}x^{-1}) - 4\tan(x) + C \\ &= \tfrac{7}{4}x^4 - \tfrac{2}{5}x^2\sqrt{x} - \frac{3}{x} - 4\tan(x) + C. \end{aligned}$$

To verify this, just differentiate $F(x)$ to recover $f(x)$.

We can also reverse the Chain Rule: we know $(\sin(3x))' = \cos(3x)\cdot(3x)' = 3\cos(3x)$, so what $F(x)$ will have $F'(x) = \cos(3x)$?

$$f(x) = \cos(3x) \qquad \Longrightarrow \qquad F(x) = \tfrac{1}{3}\sin(3x) + C.$$

On the other hand, the derivative of a product is NOT the product of derivatives (§2.3), so the antiderivative of a product is NOT the product of antiderivatives. (Similarly for quotients.) We will learn how to handle these later, but for now, we can sometimes antidifferentiate products or quotients if we can expand them into sums of Basic Antiderivatives.

$$\begin{aligned} f(x) &= \frac{x+4}{\sqrt{x}} = \frac{x}{\sqrt{x}} + \frac{4}{\sqrt{x}} = x^{1/2} + 4x^{-1/2} \\ F(x) &= \tfrac{1}{3/2}x^{3/2} + 4(\tfrac{1}{1/2}x^{1/2}) + C = \tfrac{2}{3}x\sqrt{x} + 8\sqrt{x} + C. \end{aligned}$$

EXAMPLE. Find the antiderivative of $f(x) = \sin^2(x)$. This is a product of $\sin(x)$ with itself, and we need to expand it somehow in terms of Basic Antiderivatives. A clever idea: in the identity

$$\cos(2x) \;=\; \cos^2(x) - \sin^2(x) \;=\; (1-\sin^2(x)) - \sin^2(x) \;=\; 1 - 2\sin^2(x),$$

we can solve for $\sin^2(x)$, so that:

$$\begin{aligned} f(x) &= \sin^2(x) \;=\; \tfrac{1}{2} - \tfrac{1}{2}\cos(2x), \\ F(x) &= \tfrac{1}{2}x - \tfrac{1}{2}(\tfrac{1}{2}\sin(2x)) + C \;=\; \tfrac{1}{2}x - \tfrac{1}{4}\sin(2x) + C. \end{aligned}$$

Second derivative initial value problem. Suppose a launching rocket reaches height $s(t)$ at time t. Suppose we are given the acceleration function $a(t) = t+1$, and also height and velocity at time $t = 1$, namely $s(1) = 10$ and $v(1) = 3$. We wish to find the height and velocity functions $s(t)$ and $v(t)$.

Rephrasing, we know velocity is the rate of change of position (i.e. height), $v(t) = s'(t)$, and acceleration is the rate of change of velocity, $a(t) = v'(t) = s''(t)$. Thus, we must solve the initial value problem:

$$s''(t) = t+1, \quad s(1) = 10, \quad s'(1) = 3.$$

First, we antidifferentiate $a(t) = t+1$ to get $v(t) = \frac{1}{2}t^2 + t + C$. Thus, we have:

$$v(1) = \tfrac{1}{2}(1^2) + 1 + C = 3,$$

which we can solve to get $C = \frac{3}{2}$, so that:

$$v(t) = s'(t) = \tfrac{1}{2}t^2 + t + \tfrac{3}{2}.$$

Next we antidifferentiate $v(t)$ to get:

$$s(t) \;=\; \tfrac{1}{2}(\tfrac{1}{3}t^3) + \tfrac{1}{2}t^2 + \tfrac{3}{2}t + B \;=\; \tfrac{1}{6}t^3 + \tfrac{1}{2}t^2 + \tfrac{3}{2}t + B,$$

where B is another constant (different from the previous C). Again, we can solve:

$$s(1) = \tfrac{1}{6}(1^3) + \tfrac{1}{2}(1^2) + \tfrac{3}{2}(1) + B = 10 \qquad \Longrightarrow \qquad B \;=\; \tfrac{47}{6}.$$

The final answer is:

$$s(t) \;=\; \tfrac{1}{6}t^3 + \tfrac{1}{2}t^2 + \tfrac{3}{2}t + \tfrac{47}{6}.$$

This is the unique solution, with no arbitrary constants.

Review. The derivative of $y = f(x)$ has four levels of meaning:

- Physical: If y is a quantity depending on x, the derivative $\frac{dy}{dx}|_{x=a}$ is the rate of change of y per tiny change in x away from a.
- Geometric: $f'(a)$ is the slope of the graph $y = f(x)$ near the point $(a, f(a))$, or the slope of the tangent line at that point, $y = f(a) + f'(a)(x-a)$.
- Numerical: Approximate by the difference quotient, $f'(a) \approx \frac{\Delta f}{\Delta x} = \frac{f(x)-f(a)}{x-a}$ for x near a. This gives the linear approximation $f(x) \approx f(a)+f'(a)(x-a)$.
- Algebraic: Defining $f'(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$, we prove Basic Derivatives and Derivative Rules to find $f'(x)$ for any formula $f(x)$.

Problems usually originate on the physical or geometric levels, then we translate them to the numerical or algebraic levels to solve them. For example, to find the hill tops of a given curve $y = f(x)$, a geometric problem, we consider that they must have horizontal tangents, so we take the derivative $f'(x)$ and solve for the critical points $f'(x) = 0$ algebraically, or numerically with Newton's Method.

In the previous chapter §3.9, we introduced the reverse of the derivative, the *antiderivative*. In this chapter, we will see that it has all the above levels of meaning, and connecting them will allow us to solve many new problems.

Distance problem. In §3.9, we reversed the *algebraic* derivative operation: that is, we could often recognize a given a function $f(x)$ as the derivative of some familiar function $F(x)$, obtaining an algebraic antiderivative. But this does not always work: there are many functions which are not the derivative of any formula we know, for example $f(x) = \frac{1}{x}$ or $\sqrt{x^3+1}$ or $\sin(x^2)$.*

Let us consider this on the *physical* level: if we take $f(t) = v(t)$ to be a velocity function, then the antiderivative should be the corresponding position function $F(t) = s(t)$, since velocity is the rate of change of position, $s'(t) = v(t)$. Imagine a toy car on a track which starts out at time $t = 0$ at the starting line $s(0) = 0$, and adjusts its velocity according to $v(t)$. Even if we have no algebraic formula for $s(t)$, nevertheless the car does have a position, so there must exist an antiderivative, a new function we have not imagined before.

To compute this new position function $s(t)$, we work *numerically*, taking a limit of approximations as we did in computing the derivative $\frac{ds}{dt} = \lim_{\Delta t\to 0} \frac{\Delta s}{\Delta t}$. To illustrate, we take the simple velocity function $v(t) = t^2$, and compute $s(2)$, the distance traveled from $t = 0$ to $t = 2$:

$$s'(t) \ = \ v(t) \ = \ t^2, \qquad s(0) = 0, \qquad s(2) \ = \ ??$$

Since distance = velocity × time, we can say very roughly:

$$s(2) \ \approx \ v(2)\,\Delta t \ = \ 2^2\,(2-0) \ = \ 8.$$

Notes by Peter Magyar `magyar@math.msu.edu`

* We will eventually learn that $\frac{1}{x}$ is the derivative of the logarithm $\log(x)$, but the others really are not the derivative of any formula.

This would be exact if the velocity held constant at $v(2) = 4$ for the whole time, but in fact the car starts from a standstill $v(0) = 0$, so this is a gross overestimate.

For a good approximation, we split the time interval $[0, 2]$ into $n = 20$ increments† of size $\Delta t = 0.1$, with dividing points:

$$0.0 < 0.1 < 0.2 < \cdots < 1.8 < 1.9 < 2.0.$$

We approximate a distance increment during each time increment, and add these up to get the total distance traveled; this is called a *Riemann sum*:

$$\begin{aligned} s(2) &\approx v(0.1)\Delta t + v(0.2)\Delta t + \cdots + v(1.9)\,\Delta t + v(2.0)\,\Delta t \\ &= (0.1)^2\,(0.1) + (0.2)^2\,(0.1) + \cdots + (1.9)^2\,(0.1) + (2.0)^2\,(0.1) \\ &\approx 2.9\,. \end{aligned}$$

Here we sample the velocity $v(t)$ at the end of each increment: for example, the first sample point is $t = 0.1$, the right endpoint of $t \in [0.0, 0.1]$. This is still an *overestimate* (upper Riemann sum), since the velocity is slightly less at the beginning of each increment than at the end.

To get an *underestimate* (lower Riemann sum), we should sample velocity at the beginning of each increment, where it is smallest:

$$\begin{aligned} s(2) &\approx v(0.0)\Delta t + v(0.1)\Delta t + \cdots + v(1.8)\,\Delta t + v(1.9)\,\Delta t \\ &= (0.0)^2\,(0.1) + (0.1)^2\,(0.1) + \cdots + (1.8)^2\,(0.1) + (1.9)^2\,(0.1) \\ &\approx 2.5 \end{aligned}$$

As we take more and more increments of smaller and smaller size, all estimates converge on a limiting value, which is the exact position $s(2)$.

For this simple function $v(t) = t^2$, we can compare the numerical answers with our known algebraic solution: $s(t) = \frac{1}{3}t^3$ is the unique antiderivative with $s(0) = 0$, and we have:

$$s(2) \;=\; \tfrac{1}{3}(2^3) \;=\; \tfrac{8}{3} \;\approx\; 2.66.$$

which is indeed between the lower and upper estimates above. In fact, the average of the two estimates is $\frac{2.9+2.5}{2} = 2.7$, which is the correct answer rounded to 1 decimal place.

The integral. Applied generally to any velocity $v(t)$ over any interval $t \in [0, b]$, this method specifies the value of the position $s(b)$ as a limit. We introduce a new notation for this limit, the *integral* of $v(t)$ from $t = 0$ to b:

$$s(b) \;=\; \int_0^b v(t)\,dt \;=\; \lim_{\Delta t \to 0} v(t_1)\Delta t + v(t_2)\Delta t + \cdots + v(t_n)\Delta t.$$

The integral symbol $\int$ is an elongated S standing for the *sum* of n terms; $v(t)$ stands for all the sample values $v(t_1), \ldots, v(t_n)$; and dt suggests a very small Δt for larger and larger $n \to \infty$.

† *Increment*: a small increase, a part added.

Sometimes $s(t)$ turns out to equal a known formula, sometimes it can only be computed approximately to any desired accuracy. In our example, we computed $s(2) = \int_0^2 t^2\,dt = \frac{8}{3} \approx 2.66$.

Generalizing further, suppose we are given any function $f(x)$ which we consider as the *rate of change* of an unknown function $F(x)$ for $x \in [a,b]$. Then we may compute the *total change* $F(b) - F(a)$ by the above method: that is, we compute the integral of $f(x)$ from $x = a$ to b:

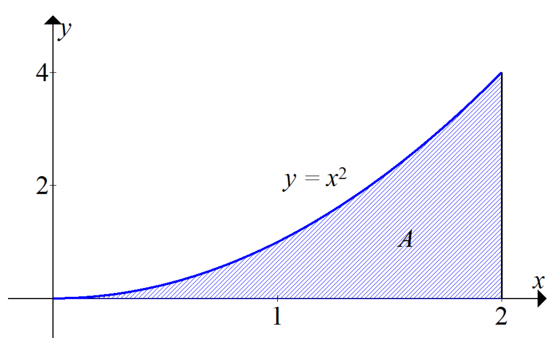
$$F(b) - F(a) \;=\; \int_a^b f(x)\,dx \;=\; \lim_{\Delta x \to 0} f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x.$$

Here we split the interval $[a,b]$ into n increments of size $\Delta x = \frac{b-a}{n}$, and choose a sample point in each increment: $x_1, x_2, \ldots, x_n$ can be the left or right endpoints, or anywhere between. Each term approximates the incremental change in $F(x)$ as the rate of change $f(x_i)$ times the length of the increment Δx. Finally, we take the limit as $n \to \infty$ and $\Delta x \to 0$.

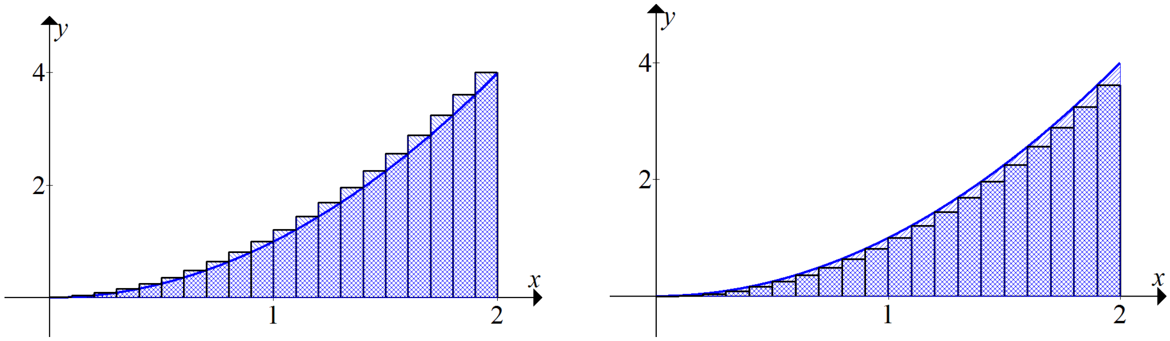
Cumulative effect of a varying influence. Another perspective: consider $f(x)$ as a varying *linear influence* on a variable $z = F(x)$, meaning that a small input increment Δx produces an output increment $\Delta z \approx f(x)\,\Delta x$. (This is the same as $f(x) \approx \frac{\Delta z}{\Delta x} \approx \frac{dz}{dx}$, i.e. $f(x)$ is the rate of change of z.) Examples: velocity influences position, acceleration influences velocity, pointwise density influences total mass, heating influences temperature, power consumption influences total energy use.

Then $\int_a^b f(x)\,dx$ computes the *cumulative effect* of this influence of $f(x)$ on z for $x \in [a,b]$, since it adds up all the Δz-increments $f(x_1)\Delta x + \cdots + f(x_n)\Delta x$.

Area problem. Now we come to one of the most surprising results in mathematics: the *geometric* interpretation of the integral. Suppose we have a function with $f(x) \geq 0$ for $x \in [a,b]$, and we wish to determine the area under the graph $y = f(x)$ and above the interval $[a,b]$ on the x-axis. For example, let us again take $f(x) = x^2$ over the interval $[0,2]$.



To approximate the area A, we cover it by 20 thin rectangles of width $\Delta x = 0.1$ (below at left):



The dividing points are again $0.0 < 0.1 < \cdots < 1.9 < 2.0$, and each rectangle reaches up to the graph at the right endpoint of an increment, giving heights $f(0.1), f(0.2), \ldots, f(2.0)$. The area A under the curve is close to the total area of the rectangles; adding up (height)×(width) for each rectangle gives:

$$A \;\approx\; f(0.1)\,\Delta x + f(0.2)\,\Delta x + \cdots + f(2.0)\,\Delta x \;\approx\; 2.9\,.$$

This is an overestimate, since the rectangles slightly overshoot the curve. To get an underestimate, we take heights at the left endpoint of each increment, fitting the rectangles under the graph (above at right):

$$A \;\approx\; f(0.0)\,\Delta x + f(0.1)\,\Delta x + \cdots + f(1.9)\,\Delta x \;\approx\; 2.5\,.$$

Clearly, this is the same computation as we did before, so it has the same answer. That is, taking the limit of thinner and thinner rectangles gives:

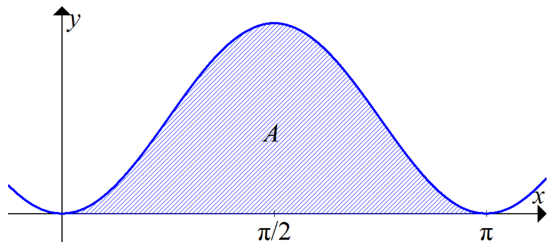
$$A \;=\; \int_0^2 x^2\,dx.$$

That is, the area under $y = x^2$ for $x \in [0, 2]$ is the *same* as the distance traveled with velocity $v(t) = t^2$ during $t \in [0, 2]$. Are you not amazed?

Why is this? Let us fix a, take $b = x$ to be a variable, and consider the area above the variable interval $[a, x]$ as a function $A(x)$. Then the rate of change of the area function is the height of the graph at the right endpoint: $A'(x) = f(x)$, since the greater the height, the taller the rightmost incremental rectangle, and the faster $A(x)$ increases. Thus, we can consider the height as a rate of change, and the area as a total change, which is just what the integral computes. (Also, we can consider the area as the cumulative effect of the graph height.)

We have seen the distance problem before in §2.7, when we used speedometer data to reconstruct odometer data, using the graph of velocity $v(t)$ to draw the graph of distance $s(t)$. We can now compute $s(t)$ more directly at a particular $t = x$ as the area under the $v(t)$ graph above the interval $t \in [0, x]$. (For consistency, we must look at the ft/sec scale for $v(t)$, not mph, since t is in sec.) This is how the $s(t)$ graphs were computed.

Approximating an integral. We compute the area A under one arch of $f(x) = \sin^2(x)$ to an accuracy of one decimal place:



We will compute:

$$A \;=\; \int_0^\pi \sin^2(x)\,dx \;\approx\; \sin^2(x_1)\,\Delta x + \sin^2(x_2)\,\Delta x + \cdots + \sin^2(x_n)\,\Delta x$$

for suitably large n, the corresponding small increment $\Delta x = \frac{b-a}{n} = \frac{\pi}{n}$, and appropriate sample points $x_1, \ldots, x_n$.

To make sure of the required accuracy, we will compute an overestimate and an underestimate (upper and lower Riemann sums). For an *overestimate*, we take $x_1, \ldots, x_n$ so that $\sin^2(x)$ is *largest* within each increment. These are not always the right endpoints, because the function is decreasing on the second half of the interval. Rather, for the increments within $[0, \frac{\pi}{2}]$, we take the right endpoints, and for the increments within $[\frac{\pi}{2}, \pi]$ we take the left endpoints. To get an *underestimate*, we take sample points where $\sin^2(x)$ is *smallest* within each increment, reversing the previous choices.

With a spreadsheet or computer algebra software, it is not difficult to take $n = 100$, $\Delta x = 0.01\pi$. The upper estimate is:

$$\sin^2(0.01\pi)\,(0.01\pi) + \sin^2(0.02\pi)\,(0.01\pi) + \cdots + \sin^2(0.50\pi)\,(0.01\pi)$$
$$+ \sin^2(0.50\pi)\,(0.01\pi) + \sin^2(0.51\pi)\,(0.01\pi) + \cdots + \sin^2(0.99\pi)\,(0.01\pi) \;\approx 1.60\,.$$

The lower estimate is:

$$\sin^2(0.00\pi)\,(0.01\pi) + \sin^2(0.01\pi)\,(0.01\pi) + \cdots + \sin^2(0.49\pi)\,(0.01\pi)$$
$$+ \sin^2(0.51\pi)\,(0.01\pi) + \sin^2(0.52\pi)\,(0.01\pi) + \cdots + \sin^2(1.00\pi)\,(0.01\pi) \;\approx 1.54\,.$$

Thus we get upper and lower bounds for our area A, and taking their average gives our best estimate:

$$1.54 < A < 1.60 \quad\Longrightarrow\quad A \;=\; 1.57 \pm 0.03\,.$$

As we have seen, the integral $F(b) = \int_a^b f(x)\,dx$ defines an antiderivative function numerically, whether or not we can find an algebraic antiderivative. But in §3.9 we were (just barely) able to find an algebraic antiderivative: $F(x) = \frac{1}{2}x - \frac{1}{4}\sin(2x)$, satisfying $F(0) = 0$. Since there can be only one such antiderivative, we find that:

$$\int_0^\pi \sin^2(x)\,dx \;=\; F(\pi) \;=\; \frac{\pi}{2} \;\approx\; 1.571\,,$$

so our numerical approximation is actually accurate to 2 decimal places.

Notation for sums. In Notes §4.1, we define the integral $\int_a^b f(x)\,dx$ as a limit of approximations. That is, we split the interval $x \in [a,b]$ into n increments of size $\Delta x = \frac{b-a}{n}$, we choose sample points $x_1, x_2, \ldots, x_n$, and we take:

$$\int_a^b f(x)\,dx \;=\; \lim_{\Delta x \to 0} f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x.$$

The sum which appears on the right is called a *Riemann sum*. Similar sums appear frequently in mathematics, and we define a special notation to handle them.

In the most general situation, we have a sequence of numbers $q_0, q_1, q_2, q_3, \ldots$ so that for any $i = 0, 1, 2, \ldots$ we have a number q_i. We consider an interval of integers $i = m, m{+}1, m{+}2, \ldots, n$, and we introduce a notation for the sum of all the q_i for $i = m$ to n:

$$\sum_{i=m}^{n} q_i \;=\; q_m + q_{m+1} + q_{m+2} + \cdots + q_n.$$

The summation symbol Σ is capital sigma, the Greek letter S, standing for "sum". The variable i is called the index of summation.

Note: In the WebWork problems, a sequence is denoted $f(i)$ instead of q_i. This is because we can consider the sequence of q_i's as a function with input i (an integer) and output q_i (a specified number).

Examples

- Letting $q_i = \sqrt{i}$, we have $q_0 = 0$, $q_1 = 1$, $q_2 = \sqrt{2}$, $q_3 = \sqrt{3}$, etc., and taking the interval of integers $i = 2, 3, 4, 5$, we have:

$$\sum_{i=2}^{5} \sqrt{i} \;=\; \sqrt{2} + \sqrt{3} + \sqrt{4} + \sqrt{5} \;\approx\; 7.38\,.$$

- Letting $q_i = 1$, we have: $\sum_{i=1}^{10} 1 = \underbrace{1 + 1 + \cdots + 1}_{10 \text{ terms}} = 10.$

- Given the sum of the first ten square numbers $1 + 4 + 9 + 16 + \cdots + 100$, we wish to write this compactly in sigma notation. Considering the terms as a sequence $q_i = i^2$, we get:

$$1 + 4 + 9 + \cdots + 100 \;=\; 1^2 + 2^2 + 3^2 + \cdots + 10^2 \;=\; \sum_{i=1}^{10} i^2\,.$$

- Given the sum of the first five odd numbers $1 + 3 + 5 + 7 + 9$, we can write this in sigma notation by considering the terms as $q_i = 2i{-}1$:

$$1 + 3 + 5 + 7 + 9 \;=\; (2(1){-}1) + (2(2){-}1) + \cdots + (2(5){-}1) \;=\; \sum_{i=1}^{5} (2i{-}1)\,.$$

Notes by Peter Magyar `magyar@math.msu.edu`

Another way would be to consider the terms as $q_i = 2i{+}1$:

$$1+3+5+7+9 \;=\; (2(0){+}1)+(2(1){+}1)+\cdots+(2(4){+}1) \;=\; \sum_{i=0}^{4}(2i{+}1)\,.$$

- The sum of the first n odd numbers, where n is an unspecified whole number, can be written as:

$$1+3+5+\cdots+(2n{-}1) \;=\; \sum_{i=1}^{n}(2i{-}1).$$

- We can write a Riemann sum as:

$$f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x \;=\; \sum_{i=1}^{n} f(x_i)\Delta x\,.$$

Summation Rules. As for limits and derivatives, we can sometimes compute summations by starting with known Basic Summations, and combining them by Summation Rules.

- *Sum:* $\displaystyle\sum_{i=m}^{n}(q_i{+}p_i) \;=\; \sum_{i=m}^{n} q_i \;+\; \sum_{i=m}^{n} p_i\,.$
- *Difference:* $\displaystyle\sum_{i=m}^{n}(q_i{-}p_i) \;=\; \sum_{i=m}^{n} q_i \;-\; \sum_{i=m}^{n} p_i\,.$
- *Constant Multiple:* $\displaystyle\sum_{i=m}^{n} C\,q_i \;=\; C\cdot\sum_{i=m}^{n} q_i$, where C does not depend on i.

Like all facts about summations, these formulas can be understood by writing out the terms in dot-dot-dot (ellipsis) notation:

$$\begin{aligned}\sum_{i=m}^{n}(q_i{+}p_i) \;&=\; (q_m{+}p_m)+(q_{m+1}{+}p_{m+1})+\cdots+(q_n{+}p_n)\\ &=\; (q_m+q_{m+1}+\cdots+q_n)+(p_m+p_{m+1}+\cdots+p_n)\\ &=\; \sum_{i=m}^{n} q_i \;+\; \sum_{i=m}^{n} p_i\,.\end{aligned}$$

Similarly for the other two rules.

Note that n is a constant not depending on i, so we may factor it out of a summation: $\sum_{i=1}^{n} ni^2 = n\sum_{i=1}^{n} i^2$. This gives a separate formula for each n: for $n=3$ it means $3(1^2){+}3(2^2){+}3(3^2) = 3(1^2{+}2^2{+}3^2)$. However, the variable i has no meaning outside the summation, and cannot be factored out: $\sum_{i=1}^{3} i2^i \overset{??}{=} i\sum_{i=1}^{3} 2^i$ is nonsense, because the left side means $1(2^1)+2(2^2)+3(2^3)$, but the right side would mean some constant "i" times $2^1{+}2^2{+}2^3$, but i is *not* a constant.

Warning: the summation of a product $\sum q_i p_i$ is NOT equal to the product of summations $(\sum q_i)(\sum p_i)$. For example: $1\cdot 1+2\cdot 2+3\cdot 3 \neq (1{+}2{+}3)(1{+}2{+}3)$.

Basic Summations. We can get some surprisingly neat formulas for certain summations:

(a) $$\sum_{i=1}^{n} 1 \;=\; n.$$

(b) $$\sum_{i=1}^{n} i \;=\; \tfrac{1}{2}n(n{+}1).$$

(c) $$\sum_{i=1}^{n} i^2 \;=\; \tfrac{1}{6}n(n{+}1)(2n{+}1).$$

Proof. (a) $\sum_{i=1}^{n} 1 = 1 + \cdots + 1$ with n terms, which indeed equals n.

(b) Taking two copies of $\sum_{i=1}^{n} i$, we can pair each term with its complement:

$$\begin{array}{rcccccccccccc}
2\cdot\sum_{i=1}^{n} i & = & & 1 & + & 2 & + & \cdots & + & n{-}1 & + & n & \\
 & & + & n & + & n{-}1 & + & \cdots & + & 2 & + & 1 & \\
\hline
 & & = & n{+}1 & + & n{+}1 & + & \cdots & + & n{+}1 & + & n{+}1 & = \; n(n{+}1).
\end{array}$$

The equation $2 \cdot \sum_{i=1}^{n} i = n(n{+}1)$, divided by 2, gives the desired formula.

(c) Consider that $(i{+}1)^3 = i^3 + 3i^2 + 3i + 1$, so that:

$$\begin{aligned}
\textstyle\sum_{i=1}^{n} (i{+}1)^3 - i^3 \;&=\; \textstyle\sum_{i=1}^{n} (3i^2 + 3i + 1) \\
&=\; 3\textstyle\sum_{i=1}^{n} i^2 + 3\sum_{i=1}^{n} i + \sum_{i=1}^{n} 1 \\
&=\; 3\textstyle\sum_{i=1}^{n} i^2 + \frac{3}{2}n(n{+}1) + n.
\end{aligned}$$

On the other hand, we have a "collapsing sum":

$$\begin{aligned}
\textstyle\sum_{i=1}^{n} (i{+}1)^3 - i^3 \;&=\; (n{+}1)^3{-}n^3 \;+\; n^3{-}(n{-}1)^3 \;+\; \cdots \;+\; 3^3{-}2^3 \;+\; 2^3{-}1^3 \\
&=\; (n{+}1)^3 - 1^3\,.
\end{aligned}$$

Solving the equation:

$$3 \cdot \sum_{i=1}^{n} i^2 + \tfrac{3}{2}n(n{+}1) + n \;=\; (n{+}1)^3 - 1$$

gives, as desired:

$$\sum_{i=1}^{n} i^2 \;=\; \tfrac{1}{3}\big((n{+}1)^3 - \tfrac{3}{2}n(n{+}1) - (n{+}1)\big) \;=\; \tfrac{1}{6}n(n{+}1)(2n{+}1)\,.$$

A similar computation will produce a formula for $\sum_{i=1}^{n} i^3$, etc.

Direct Evaluation of Integrals. We can use the above rules to simplify Riemann sums and find integrals exactly. For example, consider:

$$\int_1^3 5x\,dx \;=\; \lim_{\Delta x\to 0}\,\sum_{i=1}^{n} 5x_i\,\Delta x.$$

On the right side, we divide the interval $[1,3]$ into n increments of length $\Delta x = \frac{3-1}{n} = \frac{2}{n}$, with dividing points:

$$1 \;<\; 1{+}\Delta x \;<\; 1{+}2\Delta x \;<\; 1{+}3\Delta x \;<\; \cdots \;<\; 1{+}n\Delta x = 3.$$

In the i^{th} increment, we arbitrarily choose the sample point x_i to be the right endpoint, that is $x_i = 1 + i\,\Delta x = 1 + \frac{2}{n}i$. Thus:

$$\begin{aligned}\sum_{i=1}^{n} 5x_i\,\Delta x \;&=\; \sum_{i=1}^{n} 5\left(1+\frac{2}{n}i\right)\frac{2}{n}\\ &=\; \frac{10}{n}\sum_{i=1}^{n} 1 + \frac{20}{n^2}\sum_{i=1}^{n} i\\ &=\; \frac{10}{n}\cdot n + \frac{20}{n^2}\cdot\frac{1}{2}n(n{+}1)\\ &=\; 20 + \frac{10}{n}\,.\end{aligned}$$

(Here n is a fixed number not depending on i, such as $n = 100$ or $n = 1000$, and we can factor it out of the $\sum$.) Finally, we let $\Delta x \to 0$ or equivalently $n \to \infty$:

$$\int_1^3 5x\,dx \;=\; \lim_{\Delta x\to 0}\,\sum_{i=1}^{n} 5x_i\,\Delta x \;=\; \lim_{n\to\infty}\; 20 + \frac{10}{n} \;=\; 20\,.$$

We computed this to show in principle that Riemann sums can be evaluated directly, but this is far from the easiest way to compute an integral. Geometrically, the integral equals the trapezoid area below the graph $y = 5x$ and above the interval $[1,3]$ on the x-axis. Since (trapezoid area) = (width)×(average height), we get that the integral is $A = (3{-}1)(\frac{5(1)+5(3)}{2}) = 20$.

Physically, if $v(t) = 5t$ is a velocity, then the integral $\int_1^3 v(t)\,dt$ is the distance traveled from $t = 1$ to $t = 3$. Since the position $s(t)$ is an antiderivative, we must have $s(t) = \frac{5}{2}t^2 + C$, so the distance traveled is $s(3) - s(1) = \frac{5}{2}(3^2) - \frac{5}{2}(1^2) = 20$.

Precise definition. We have defined the integral $\int_a^b f(x)\,dx$ as a number approximated by Riemann sums. The integral is useful because, given a velocity function, it computes distance traveled; given a graph, it computes an area between the graph and the x-axis. More generally, given a varying rate of change, the integral computes the total change; or given a varying linear influence, the integral computes its cumulative effect.

The parts of the notation $\int_a^b f(x)\,dx$ have their own names: $\int$ is the integral sign; a is the *lower limit of integration*;* b is the *upper limit of integration*; $f(x)$ is the *integrand*; and x is the *variable of integration*. Note that the variable of integration is named only for convenience, and changing it does not change the value of the integral: $\int_a^b f(x)\,dx = \int_a^b f(t)\,dt = \int_a^b f(r)\,dr$.

Now we give the formal definition of integrals on the numerical level, as we did for limits in §1.7 and derivatives in §2.1.

Definition: Given a function $f(x)$ and numbers $a \le b$.

- For each positive integer n, we divide the interval $x \in [a,b]$ into n increments of width $\Delta x = \frac{b-a}{n}$, with division points:

$$a \;<\; a{+}\Delta x \;<\; a{+}2\Delta x \;<\; \cdots \;<\; a{+}n\Delta x = b,$$

  and we choose sample points $x_1, \ldots, x_n$ with x_i anywhere in the i^{th} increment: $a{+}(i{-}1)\Delta x \le x_i \le a{+}i\Delta x$. Then we let:

$$\int_a^b f(x)\,dx \;=\; \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\Delta x \;=\; \lim_{n\to\infty} f(x_1)\Delta x{+}\cdots{+}f(x_n)\Delta x.$$

- The function $f(x)$ is *integrable* over $[a,b]$ whenever the above limit exists for every possible choice of sample points x_i.†

Integrable functions. Most functions are integrable unless they have a vertical asymptote. To be precise:

Theorem: Assume $f(x)$ is continuous for all $x \in [a,b]$, except possibly at a finite list of removable or jump discontinuities (see §1.8).

Then $f(x)$ is integrable, meaning its Riemann sums converge to a well-defined limit $L = \int_a^b f(x)\,dx$ for any choice of sample points.

This is proved in courses on Real Analysis.

To understand integrability better, let's examine the non-integrable function $f(x) = \frac{1}{x^2}$ on $[a,b] = [0,1]$. (We arbitrarily set $f(0) = 0$ to make $f(x)$ defined for

---

Notes by Peter Magyar `magyar@math.msu.edu`

* Here we use "limit" to mean a boundary, *not* a value approached by approximations.

† Even more formally, $\int_a^b f(x)\,dx = L$ means that for any error tolerance $\varepsilon > 0$, there is some lower bound N such that any Riemann sum with more than N terms is forced close to L within an error of ε: that is, $n > N$ forces $\left|\sum_{i=1}^n f(x_i)\Delta x - L\right| \;<\; \varepsilon$.

all $x \in [0,1]$.) The function has a vertical asymptote discontinuity at $x=0$, so the Theorem does not apply. If we attempt to compute $\int_0^1 \frac{1}{x^2}\,dx$ by a Riemann sum with $\Delta x = \frac{1-0}{n} = \frac{1}{n}$ and sample points $x_i = a+i\Delta x = \frac{i}{n}$, we get:

$$\sum_{i=1}^{n} \frac{1}{x_i^2}\cdot\Delta x \;=\; \sum_{i=1}^{n} \frac{n^2}{i^2}\cdot\frac{1}{n} \;=\; \sum_{i=1}^{n} n\cdot\frac{1}{i^2} \;=\; n+\tfrac{n}{4}+\cdots \;>\; n\,.$$

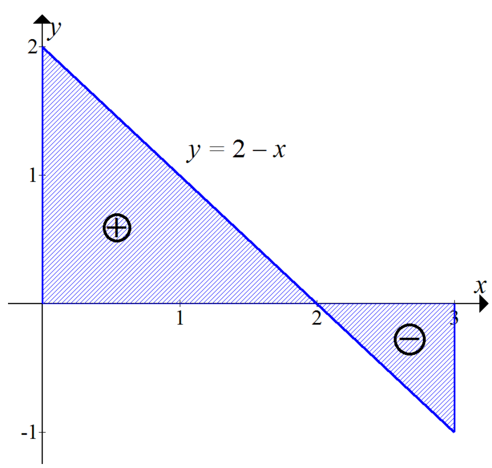
Thus, as $n\to\infty$, the Riemann sum also gets larger and larger, and does not approach a finite limit. Geometrically, this means there is an infinite area under the curve and above the interval $[0,1]$ on the x-axis.

Negative integrand. So far, we have considered $\int_a^b f(x)\,dx$ with positive integrand $f(x)\geq 0$, in which case the integral is a *positive* number (in fact, an area). Now suppose $f(x)\leq 0$: our definition of the integral still makes sense, but it gives a *negative* number. For example, for the constant function $f(x)=-1$, we have:

$$\begin{aligned}\int_1^3(-1)\,dx \;&=\; \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i)\Delta x \;=\; \lim_{n\to\infty}\sum_{i=1}^{n}(-1)(\tfrac{3-1}{n})\\ &=\; \lim_{n\to\infty}(\tfrac{-2}{n})\sum_{i=1}^{n}1 \;=\; \lim_{n\to\infty}(\tfrac{-2}{n})\cdot n \;=\; -2\,.\end{aligned}$$

Geometrically, we think of each term $f(x_i)\,\Delta x$ as (height)×(width) with a "negative height" $f(x_i)\leq 0$, and we count this as a "negative area". For a general graph $y=f(x)$ passing above and below the x-axis, $\int_a^b f(x)\,dx$ computes the "signed area" between the graph and the interval $[a,b]$ on the x-axis, with regions above the x-axis counting as positive area, and regions below counting negative.

EXAMPLE: We could evaluate the integral $\int_0^3(2-x)\,dx$ with Riemann sums as above, but it is easier geometrically. The function $f(x)=2-x$ has x-intercept $x=2$; it is positive for $x\in[0,2]$ and negative for $x\in[2,3]$. Thus the integral is the area of the triangle above $[0,2]$, *minus* the area of the triangle below $[2,3]$:



But (triangle area) $=\frac{1}{2}$(base)×(height), so $\int_0^3(2-x)\,dx = \frac{1}{2}(2)(2)-\frac{1}{2}(1)(1)=\frac{3}{2}$.

Reversing limits of integration. If we take the limits of integration to be the same, $a = b$, then $\Delta x = \frac{b-a}{n} = 0$, so every Riemann sum is zero, and we get:

$$\int_a^a f(x)\,dx \;=\; 0\,.$$

This is clear geometrically, since the area above a one-point interval $[a, a]$ is zero.

Next, we give a meaning to switching the two limits of integration, by defining:

$$\int_b^a f(x)\,dx \;=\; -\int_a^b f(x)\,dx\,.$$

Geometrically, in $\int_b^a f(x)\,dx$ with $a < b$, we imagine x running backward from b to a with a negative increment $\Delta x = \frac{a-b}{n} < 0$. Since each term $f(x_i)\Delta x$ has "negative width" Δx, the integral becomes the negative of $\int_a^b f(x)\,dx$. If $f(x_i)$ is also negative, then both width and height are negative and the integral is positive: for example, $\int_3^1 (-1)\,dx = -\int_1^3 (-1)\,dx = -(-2) = 2$.

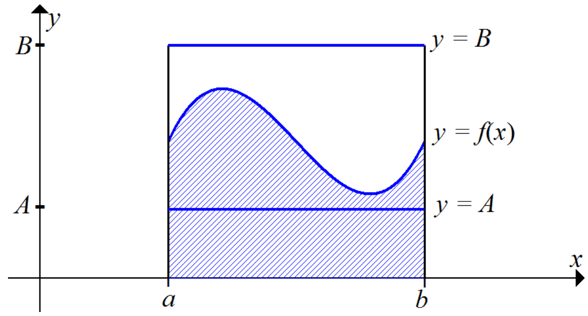
We bother with this definition only so as to improve the Splitting Rule below.

Integral Rules. At the end of §4.1 Part 2 we gave a direct computation of an integral as a limit of Riemann sums. In later sections, we will find much better algebraic methods to compute integrals using the Fundamental Theorems of Calculus. For now, we will rely on a few Basic Integrals, and the following Rules to combine them. Let a, b, c be any points on the x-axis; $f(x), g(x)$ any integrable functions; and A, B, C any constants. Then we have:

- *Sum:* $\displaystyle\int_a^b f(x)+g(x)\,dx \;=\; \int_a^b f(x)\,dx \,+\, \int_a^b g(x)\,dx\,.$
- *Difference:* $\displaystyle\int_a^b f(x)-g(x)\,dx \;=\; \int_a^b f(x)\,dx \,-\, \int_a^b g(x)\,dx\,.$
- *Constant Multiple:* $\displaystyle\int_a^b C\,f(x)\,dx \;=\; C\cdot\int_a^b f(x)\,dx\,.$
- *Domination:* If $f(x) \le g(x)$, then: $\displaystyle\int_a^b f(x)\,dx \;\le\; \int_a^b g(x)\,dx\,.$
- *Bounds:* If $A \le f(x) \le B$, then: $\displaystyle(b-a)A \;\le\; \int_a^b f(x)\,dx \;\le\; (b-a)B\,.$
- *Splitting:* $\displaystyle\int_a^c f(x)\,dx \;=\; \int_a^b f(x)\,dx \,+\, \int_b^c f(x)\,dx\,.$

Proof. The Sum, Difference, and Constant Multiple Rules follow directly from the corresponding rules for summations in §4.1 Part 2, applied to Riemann sums.

The Bounds Rule makes sense geometrically because $0 \le A \le f(x) \le B$ means the graph $y = f(x)$ is above the line $y = A$ and below $y = B$. Thus the area $\int_a^b f(x)\,dx$ below $y = f(x)$ and above $[a, b]$ contains a rectangle with (width)$\times$(height) $= (b-a)A$, and the area is contained inside a rectangle with (width)$\times$(height) $= (b-a)B$.



Computing formally, $A \le f(x_i) \le B$ implies $\sum_{i=1}^n A\,\Delta x \le \sum_{i=1}^n f(x_i)\Delta x \le \sum_{i=1}^n B\,\Delta x$. Hence:

$$\sum_{i=1}^n A\,\Delta x \;=\; \sum_{i=1}^n A\cdot\tfrac{b-a}{n} \;=\; \tfrac{(b-a)A}{n}\sum_{i=1}^n 1 \;=\; \tfrac{(b-a)A}{n}\cdot n \;=\; (b-a)A,$$

and similarly for the upper bound. Taking limits as $n\to\infty$ gives the desired inequalities. The Domination Rule is similar.

The Splitting Rule is intuitive when $a \le b \le c$. The interval $[a,c]$ splits as the union of two sub-intervals, $[a,b]\cup[b,c]$, so the area above $[a,c]$ is the sum of the areas above $[a,b]$ and $[b,c]$, i.e. $\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx$.

Furthermore, because of our extended definition of integrals, the Splitting Rule is valid no matter what the relative positions of a, b, c. For example, if $a \le c \le b$, then $[a,b] = [a,c]\cup[c,b]$ and clearly $\int_a^c f(x)\,dx + \int_c^b f(x)\,dx \;=\; \int_a^b f(x)\,dx$. Moving $\int_c^b$ to the other side, we get:

$$\int_a^c f(x)\,dx \;=\; \int_a^b f(x)\,dx - \int_c^b f(x)\,dx \;=\; \int_a^b f(x)\,dx + \int_b^c f(x)\,dx\,,$$

so the very same Splitting Rule applies: $\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx$.

Another example: if $a = c$, the Splitting Rule says:

$$\int_a^a f(x)\,dx \;=\; \int_a^b f(x)\,dx + \int_b^a f(x)\,dx,$$

which is true since both sides are zero.

Basic Integrals:

$$\int_a^b 1\,dx \;=\; b-a, \qquad \int_a^b x\,dx \;=\; \tfrac{1}{2}b^2 - \tfrac{1}{2}a^2, \qquad \int_a^b x^2\,dx \;=\; \tfrac{1}{3}b^3 - \tfrac{1}{3}a^3.$$

Later, we will easily evaluate these integrals by the Fundamental Theorems. For now, we can prove them directly from the Basic Summations in §4.1 Part 2. For the third and hardest formula, we take increment $\Delta x = \frac{b-a}{n}$, sample points $x_i = a + i\Delta x$, and $f(x_i) = (a+i\Delta x)^2 = a^2 + 2ai\,\Delta x + i^2(\Delta x)^2$, giving Riemann sum:

$$\begin{aligned}
\sum_{i=1}^{n} f(x_i)\Delta x &= \sum_{i=1}^{n} a^2\Delta x + 2ai(\Delta x)^2 + i^2(\Delta x)^3 \\
&= a^2\Delta x \sum_{i=1}^{n} 1 + 2a(\Delta x)^2 \sum_{i=1}^{n} i + (\Delta x)^3 \sum_{i=1}^{n} i^2 \\
&= a^2 \tfrac{(b-a)}{n}\, n + 2a\tfrac{(b-a)^2}{n^2}\, \tfrac{1}{2}n(n+1) + \tfrac{(b-a)^3}{n^3}\, \tfrac{1}{6}n(n+1)(2n+1) \\
&= a^2(b-a) + 2a(b-a)^2(\tfrac{1}{2}+\tfrac{1}{2n}) + (b-a)^3(\tfrac{1}{3}+\tfrac{1}{2n}+\tfrac{1}{6n^2})
\end{aligned}$$

Taking the $\lim_{n\to\infty}$, the terms with n in the denominator disappear, and we get:

$$\int_a^b x^2\,dx \;=\; a^2(b-a) + 2a(b-a)^2(\tfrac{1}{2}) + (b-a)^3(\tfrac{1}{3}) \;=\; \tfrac{1}{3}b^3 - \tfrac{1}{3}a^3.$$

Examples.

- Evaluate the integral:

  $$\int_{-2}^{7} (3t-5)^2\,dt.$$

  We use the Integral Rules to reduce the problem to Basic Integrals. Since $\int f(x)^2\,dx$ is NOT equal to $(\int f(x)\,dx)^2$, we must expand the integrand and apply the Sum, Difference, and Constant Multiple Rules:

  $$\begin{aligned}
  \textstyle\int_{-2}^{7}(3t-5)^2\,dt &= \textstyle\int_{-2}^{7}(9t^2 - 30t + 25)\,dt \\
  &= 9\textstyle\int_{-2}^{7} t^2\,dt \;-\; 30\int_{-2}^{7} t\,dt \;+\; 25\int_{-2}^{7} 1\,dt \\
  &= 9\left(\tfrac{1}{3}7^3 - \tfrac{1}{3}(-2)^3\right) - 30\left(\tfrac{1}{2}7^2 - \tfrac{1}{2}(-2)^2\right) + 25\,(7-(-2)) \\
  &= 603\,.
  \end{aligned}$$

  Note that the variable of integration (t or x) is irrelevant.

- Find an upper bound for:

  $$\int_{-2}^{7} (3x-5)^2\,(1+\cos(x^2))\,dx.$$

  That is, we do not ask for an exact value, only an overestimate. We know that $1 + \cos(x^2) \le 2$, so the Domination Rule gives:

  $$\textstyle\int_{-2}^{7}(3x-5)^2(1+\cos(x^2))\,dx \;\le\; \int_{-2}^{7}(3x-5)^2(2)\,dx \;=\; 2(603) \;=\; 1206\,.$$

Integral as antiderivative. In §4.1, we were given a velocity function $v(t)$, and we wanted to determine the distance traveled over a given time interval $t \in [a,b]$. The answer was an integral defined as a Riemann sum, adding up (velocity)×(time) over many small time increments of length Δt:

$$\text{distance traveled} \;=\; s(b)-s(a) \;=\; \lim_{\Delta t\to 0}\sum_{i=1}^{n} v(t_i)\,\Delta t \;=\; \int_a^b v(t)\,dt.$$

Assuming initial position $s(a)=0$, and taking $b=x$, a variable endpoint, this means:*

$$s(x) \;=\; \int_a^x v(t)\,dt.$$

Since the rate of change of position is velocity, $s'(x)=v(x)$, this always computes an antiderivative function for $v(t)$, even if it is impossible to get an antiderivative algebraically by reversing differentiation formulas.

First Fundamental Theorem. Stating the above formally:

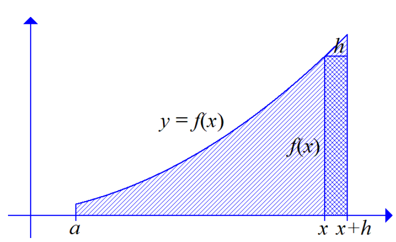
> *Theorem:* Let $f(x)$ be continuous for all $x \in [a,b]$ and define the function:†
>
> $$I(x) \;=\; \int_a^x f(t)\,dt.$$
>
> Then $I'(x)=f(x)$ for $x\in(a,b)$, and $I(x)$ is the unique antiderivative of $f(x)$ with $I(a)=0$.

In more general physical terms: the rate of change of a cumulative effect up to some time is the strength of the effect at that time.

Proof. In a rigorous argument, we cannot use our physical intuition about velocity and position, and we do not even know if there exists any anti-derivative function. Rather, we define the candidate anti-derivative: $I(x)=\int_a^x f(x)\,dx$, and we compute its derivative from the definition: $I'(x)=\lim\limits_{h\to 0}\frac{I(x+h)-I(x)}{h}$. We have:

$$\frac{I(x+h)-I(x)}{h} \;=\; \frac{1}{h}\left(\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt\right) \;=\; \frac{1}{h}\int_x^{x+h} f(t)\,dt\,,$$

since $\int_a^{x+h} = \int_a^x + \int_x^{x+h}$ for all h (even $h<0$).



Notes by Peter Magyar `magyar@math.msu.edu`

*We use the new variable x to avoid $s(t) \overset{??}{=} \int_a^t v(t)\,dt$, which would imply nonsense like $s(2) \overset{??}{=} \int_a^2 v(2)\,d2$.

†Again, we must use different letters for the limit of integration x and the variable of integration t.

Geometrically, we see that if h is small enough, the region above $[x, x{+}h]$ is approximately a rectangle with height $f(x)$ and width h, so $\int_x^{x+h} f(x)\,dx \approx f(x)h$, and:

$$I'(x) \;\approx\; \frac{I(x{+}h) - I(x)}{h} \;=\; \frac{1}{h}\int_x^{x+h} f(t)\,dt \;\approx\; \tfrac{1}{h}(f(x)h) \;=\; f(x),$$

with approximations turning into equalities as $h \to 0$, as claimed by the Theorem.

However, geometric inspection is also insufficient for a proof, because any picture only shows a particular case, and is not numerically precise. To control errors, we take the absolute minimum value N and the absolute maximum value M of the continuous function $f(x)$ on $[x, x{+}h]$, using the Extremal Value Theorem (§3.1).[‡] (To indicate that these depend on h, we write N_h, M_h.) Now, $N_h \le f(t) \le M_h$ for $t \in [x, x{+}h]$, so by the Bounds Rule for integrals (§4.2) we have:

$$((x{+}h){-}x)N_h \;\le\; \int_x^{x+h} f(t)\,dt \;\le\; ((x{+}h){-}x)M_h \quad\Longrightarrow\quad N_h \;\le\; \frac{1}{h}\int_x^{x+h} f(t)\,dt \;\le\; M_h.$$

As h gets very small, the interval $[x, x{+}h]$ gets closer and closer to the single point x, and the absolute minimum and maximum over this tiny interval must approach $f(x)$ by continuity: that is, $\lim_{h\to 0} N_h = \lim_{h\to 0} M_h = f(x)$. Also, by the above we have:

$$N_h \;\le\; \frac{I(x{+}h) - I(x)}{h} \;=\; \frac{1}{h}\int_x^{x+h} f(t)\,dt \;\le M_h\,.$$

Applying the Squeeze Theorem for limits (§1.6), we find what we wanted:

$$I'(x) \;=\; \lim_{h\to 0} \frac{I(x{+}h) - I(x)}{h} \;=\; \lim_{h\to 0} N_h \;=\; \lim_{h\to 0} M_h \;=\; f(x)\,,$$

As for the uniqueness part of the conclusion, it is clear that $I(a) = \int_a^a f(t)\,dt = 0$, and there is a unique antiderivative with this initial value by the Antiderivative Theorem (§3.9), which is a version of the Uniqueness Theorem (§3.2). Note how we have used almost all of our previous theory in proving this culminating Theorem.

Derivative of integral functions. The above Theorem can be stated as a Basic Derivative formula for $I(x) = \int_a^x f(t)\,dt$, where $f(t)$ is continuous:

$$I'(x) \;=\; \frac{d}{dx}\left(\int_a^x f(t)\,dt\right) \;=\; f(x)\,.$$

Here a is any constant, x is the input variable, and t is a "dummy variable" which only has meaning inside the integral.

For another function $g(x)$, we can take its composition with $I(x)$. Then the above Basic Derivative together with the Chain Rule (§2.5) implies:

$$I(g(x))' \;=\; \frac{d}{dx}\left(\int_a^{g(x)} f(t)\,dt\right) \;=\; I'(g(x))\cdot g'(x) \;=\; f(g(x))\cdot g'(x)\,.$$

[‡]Here we assume $h > 0$. The case $h < 0$ is the same except for a few sign changes.

EXAMPLE: Find the derivative of $F(x) = \int_{2x}^{x^3} \sin(x)\,dx$. We have:

$$\begin{aligned} F'(x) &= \frac{d}{dx}\left(\int_{2x}^{x^3} \sin(x)\,dx\right) = \frac{d}{dx}\left(\int_0^{x^3} \sin(x)\,dx - \int_0^{2x} \sin(x)\,dx\right) \\ &= \sin(x^3)\cdot(x^3)' - \sin(2x)\cdot(2x)' = 3x^2\sin(x^3) - 2\sin(2x)\,. \end{aligned}$$

EXAMPLE: Find the derivative of $F(x) = \int_{2a}^{b^3} \sin(t)\,dt$. Here a, b are constants, and hence so are $2a, b^3$. In fact, the right hand side does not depend on the variable x, and is a constant function with derivative $F'(x) = 0$! This also follows from the Chain Rule, since $\sin(2a)\cdot 2(a)' = 0$ and $\sin(b^3)\cdot(b^3)' = \sin(b^3)\cdot 3b^2\cdot(b)' = 0$.

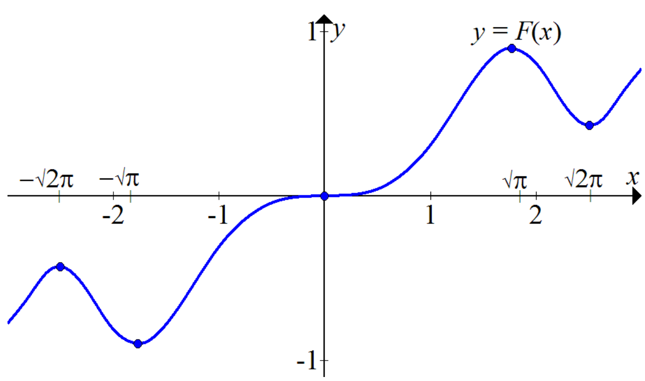
Sketching integral functions. Since an antiderivative $I(x) = \int_a^x f(t)\,dt$ might be a completely new function for which no elementary formula is possible, it might seem mysterious. However, we can find its values numerically with sufficient accuracy by computing Riemann sums on a spreadsheet, and plot these to get a good idea of the graph.

A geometric strategy is to use the derivative $I'(x) = f(x)$ for sketching $y = I(x)$, as in §3.3 and §3.5. That is, the *slope* of the graph $y = I(x)$ is given by the *height* of $y = f(x)$.

EXAMPLE: Graph the function $I(x) = \int_0^x \sin(t^2)\,dt$. The critical points of $I(x)$ are solutions of $I'(x) = 0$ or undefined, i.e. $f(x) = \sin(x^2) = 0$ (defined for all x). This happens when $x^2 = 2k\pi$ for any integer k, so the critical points are $x = 0, \pm\sqrt{\pi}, \pm\sqrt{2\pi}, \ldots$. Sign chart:

x		$-\sqrt{2\pi}$		$-\sqrt{\pi}$		0		$\sqrt{\pi}$		$\sqrt{2\pi}$	
$I'(x)$	$+$	0	$-$	0	$+$	0	$+$	0	$-$	0	$+$
$I(x)$	$\nearrow$	-0.43	$\searrow$	-0.89	$\nearrow$	0	$\nearrow$	0.89	$\searrow$	0.43	$\nearrow$

For inflection points, we solve $I''(x) = 0$, i.e. $f'(x) = 2x\cos(x^2) = 0$, so $x = 0, \pm\sqrt{\frac{\pi}{2}}$, $\pm\sqrt{\frac{3\pi}{2}}, \ldots$. Thus, the general shape of the graph is clear, and we can get specific points $(b, I(b))$ from computing a Riemann sum for $\int_0^b \sin(t^2)\,dt$.



From the 180° rotational symmetry of the graph, it looks like $I(x)$ is an odd function,

$I(-x) = -I(x)$. This is because $f(x) = \sin(x^2)$ is an even function, $f(-x) = f(x)$, so:

$$\begin{aligned} I(-b) &= \textstyle\int_0^{-b} \sin(t^2)\,dt \;=\; -\int_{-b}^0 \sin(t^2)\,dt \\ &= -(\text{area under } y = \sin(x^2) \text{ above } x \in [-b, 0]) \\ &= -(\text{area under } y = \sin(x^2) \text{ above } x \in [0, b]) \\ &= -\textstyle\int_0^b \sin(t^2)\,dt \;=\; -I(b) \end{aligned}$$

Second Fundamental Theorem. This is a trick to easily evaluate many integrals, which we already used to find some exact values in §4.1.

> *Theorem:* Suppose $F(x)$ is some known antiderivative with $F'(x) = f(x)$. Then:
>
> $$\int_a^b f(x)\,dx \;=\; F(b) - F(a)\,.$$
>
> That is, if $f(x)$ is the rate of change of $F(x)$, then the integral $\int_a^b f(x)\,dx$ is the total change of $F(x)$ from $x = a$ to b.

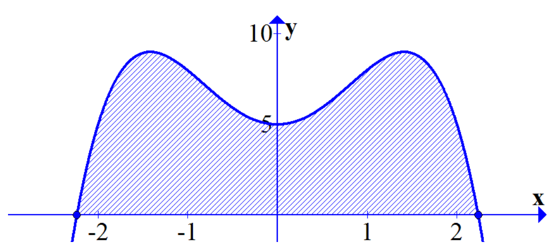
Put another way: the cumulative effect of a rate of change is a total change.

Proof. Since $F(x)$ is a particular antiderivative of $f(x)$, the Uniqueness Theorem (§3.9, §3.2) says that the general antiderivative is $F(x) + C$ for any constant C. But the First Fundamental Theorem says the integral function $I(x) = \int_0^x f(t)\,dt$ is also an antiderivative of $f(x)$, so we must have $I(x) = F(x) + C$. Since we know the initial condition $I(a) = \int_a^a f(t)\,dt = 0$, we get $I(a) = F(a) + C = 0$, and $C = -F(a)$. Therefore $I(x) = F(x) - F(a)$ and $\int_a^b f(x)\,dx = I(b) = F(b) - F(a)$ as desired.[§]

EXAMPLE: Evaluate the integral: $\int_{-\sqrt{5}}^{\sqrt{5}} 5{+}4x^2{-}x^4\,dx$. Reversing our Derivative Rules as we did in §3.9, we see that $F(x) = 5x{+}\frac{4}{3}x^3{-}\frac{1}{5}x^5$ is an antiderivative. By the Theorem:

$$\int_{-\sqrt{5}}^{\sqrt{5}} 5{+}4x^2{-}x^4\,dx \;=\; F(\sqrt{5}) - F(-\sqrt{5}) \;=\; \tfrac{20}{3}\sqrt{5} - (-\tfrac{20}{3}\sqrt{5}) \;=\; \tfrac{40}{3}\sqrt{5} \;\approx\; 29.81$$

EXAMPLE: Determine the area under the curve $y = 5{+}4x^2{-}x^4$ and above the x-axis.



We must find the limits of integration, which are the x-intercepts of the graph. Subtituting $u = x^2$, the equation becomes $5 + 4u - u^2 = 0$, which we can solve by the Quadratic Formula as $u = -1$ or 5, so $x = \pm\sqrt{u} = \pm\sqrt{5}$. Thus the area is $\int_{-\sqrt{5}}^{\sqrt{5}} 5{+}4x^2{-}x^4\,dx = \frac{40}{3}\sqrt{5}$. (Check: Graph's average height ≈ 7, base ≈ 4, so area ≈ 28, agreeing with above.)

[§] The variable of integration, x or t, is irrelevant, provided it doesn't conflict with the limits of integration.

Review. The integral $\int_a^b f(x)\,dx$ has four levels of meaning.

- Physical: Suppose y, z are physical variables determined as continuous functions of an independent variable x, so that $y = f(x)$ and $z = F(x)$. If y is the rate of change of z, i.e. $y = \frac{dz}{dx}$, then the integral of y is the cumulative total change of z between $x = a$ and $x = b$. In Leibnitz notation:*

$$\int_a^b y\;dx \;=\; z\Big|_{x=a}^{x=b}\;.$$

In Newton notation, $f(x) = F'(x)$, and:

$$\int_a^b f(x)\,dx \;=\; F(b) - F(a)\,.$$

This is the *Second Fundamental Theorem of Calculus.*

If we know an initial value $F(a)$, we have $F(x) = F(a) + \int_a^x f(t)\,dt$, and:

$$F'(x) \;=\; \frac{d}{dx}\int_a^x f(t)\,dt \;=\; f(x)\,.$$

This is the *First Fundamental Theorem of Calculus.*

- Geometric: The integral is the area between the graph $y = f(x)$ and the interval $x \in [a, b]$, counting area above the x-axis as positive, area below the x-axis as negative.

- Numerical: To compute the integral, we divide $x \in [a, b]$ into n increments of width $\Delta x = \frac{b-a}{n}$, and choose sample points $x_1, \ldots, x_n$, one in each increment. Then the integral is approximated by a Riemann sum:

$$\int_a^b f(x)\,dx \;\approx\; \sum_{i=1}^{n} f(x_i)\Delta x \;=\; f(x_1)\Delta x + \cdots + f(x_n)\Delta x\,.$$

The exact integral is the limit of these approximations as $n \to \infty$, $\Delta x \to 0$.

- Algebraic: If, by reversing derivative formulas, we can find a formula for an antiderivative $F(x)$ with $F'(x) = f(x)$, then we can compute $\int_a^b f(x)\,dx = F(b) - F(a)$ by the Second Fundamental Theorem. Later, we will develop techniques for finding $F(x)$, such as the Substitution Method which reverses the Chain Rule.

Notes by Peter Magyar `magyar@math.msu.edu`

* In Leibnitz notation, a function is denoted by its output variable, such as $z = F(x)$, and its derivative function is $\frac{dz}{dx} = F'(x)$. A particular output value of the function is denoted: $z|_{x=a} = F(a)$; and the change in the value over an interval $x \in [a, b]$ is denoted: $z|_{x=a}^{x=b} = F(b) - F(a)$.

Indefinite integral notation. Since antiderivatives are so closely related to integrals by the Fundamental Theorems, we adopt the integral sign as a notation for the most general antiderivative of a function:

$$\int f(x)\,dx \;=\; F(x) + C \quad \text{for all } C.$$

Here $F(x)$ is a particular antiderivative: $F'(x) = f(x)$; and $F(x) + C$ means the family of all antiderivatives, one for every constant C (§3.9). This family is called the *indefinite integral*, with no specific limits of integration next to the $\int$ sign.

EXAMPLE: Since $\frac{d}{dx}(x^3) = 3x^2$ and $\frac{d}{dx}(\frac{1}{3}x^3) = x^2$, we have the indefinite integral:

$$\int x^2\,dx \;=\; \tfrac{1}{3}x^3 + C\,.$$

EXAMPLE: Suppose a car with position function $s(t)$ lurches forward with velocity $v(t) = 10t + 10\sin(\pi t)$ m/sec. How far does it travel from $t = 0$ to $t = 3$ sec? Making use of the antiderivative table in §3.9, we first find the indefinite integral:

$$\int 10t + 10\sin(\pi t)\,dt \;=\; 5t^2 - \tfrac{10}{\pi}\cos(\pi t) + C$$

Since velocity is the rate of change of position, the total change in position is the definite integral:

$$s(3) - s(1) \;=\; \int_0^3 10t + 10\sin(\pi t)\,dt \;=\; 5t^2 - \tfrac{10}{\pi}\cos(\pi t)\,\Big|_{t=0}^{t=3}$$

$$=\; \big(5(3^2) - \tfrac{10}{\pi}\cos(\pi{\cdot}3)\big) - \big(5(0^2) - \tfrac{10}{\pi}\cos(\pi{\cdot}0)\big) \;=\; 45 + \tfrac{20}{\pi} \;\approx\; 51.4 \text{ meters.}$$

Average of a function. In the numerical definition of integral above, we can rewrite the Riemann sum as:

$$\sum_{i=1}^{n} f(x_i)\,\Delta x \;=\; \sum_{i=1}^{n} f(x_i)\,\frac{b{-}a}{n} \;=\; (b{-}a)\,\frac{f(x_1) + \cdots + f(x_n)}{n}\,.$$

This is just the interval length $(b{-}a)$ times the average of the sample values $f(x_1), \ldots, f(x_n)$. The integral is the limit of this as $n \to \infty$, which becomes $(b{-}a)$ times the average of a more and more dense set of sample values:

$$\int_a^b f(x)\,dx \;=\; (b{-}a)\,\lim_{n\to\infty} \frac{f(x_1) + \cdots + f(x_n)}{n}\,.$$

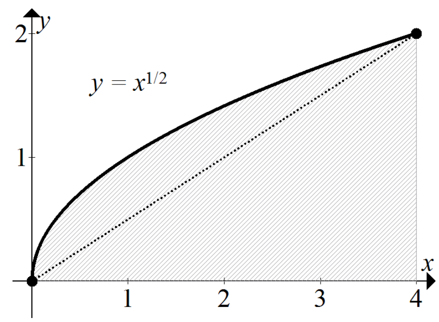
We define the average of $f(x)$ over all $x \in [a, b]$ to be the above limit. Hence:

$$\text{Average of } f(x) \text{ over } [a, b] \;=\; f_{\text{ave}} \;=\; \frac{1}{b{-}a}\int_a^b f(x)\,dx\,.$$

EXAMPLE: Find the average value of $f(x) = \sqrt{x}$ over $x \in [0,4]$. The indefinite integral is: $\int \sqrt{x}\, dx = \int x^{1/2}\, dx = \frac{2}{3}x^{3/2} + C$. The average is thus:

$$\frac{1}{4{-}0}\int_0^4 \sqrt{x}\, dx \;=\; \frac{1}{4}\,\frac{2}{3}x^{\frac{3}{2}}\Big|_{x=0}^{x=4} \;=\; \frac{4}{3} \;\approx\; 1.3\ .$$

That is, the function varies between $0 \le \sqrt{x} \le 2$ over the interval, but its average value is higher than the halfway point 1.0, because the graph bulges above the straight line from $(0,0)$ to $(4,2)$.



A geometric way to picture the average of a positive $f(x)$ over $[a,b]$ is to think of the area under the curve as a fluid. If we remove the curve and contain the fluid between the walls $x = a$ and $x = b$, then the level of the fluid is the average of the function. In the picture, the fluid under the straight line would fill the container between $x = 0$ and $x = 4$ to the midpoint level $y = 1$; but with extra fluid under $y = \sqrt{x}$ and above the line, the average of $f(x) = \sqrt{x}$ is higher: $y = \frac{4}{3}$.

EXAMPLE: We first discussed average velocity §1.4 as distance traveled divided by time elapsed, then defined instantaneous velocity as a limit of this, leading to the definition of derivative. Having made this definition, we can start with a position function $s(t)$ and its velocity function $s'(t) = v(t)$, and find over time interval $t \in [a,b]$:

$$v_{\text{ave}} \;=\; \frac{1}{b{-}a}\int_a^b v(t)\, dt \;=\; \frac{1}{b{-}a}\int_a^b s'(t)\, dt \;=\; \frac{s(b){-}s(a)}{b{-}a}$$

by the Second Fundamental Theorem (§4.3). That is, the average of the velocity function is indeed the distance traveled $s(b){-}s(a)$, divided by the time elapsed $b{-}a$.†

Mean Value Theorem for Integrals

> If $f(x)$ is continuous for $x \in [a,b]$, then there is some $c \in (a,b)$ where $f(c)$ equals the average of $f(x)$ over the interval: $f(c) = f_{\text{ave}} = \frac{1}{b{-}a}\int_a^b f(x)\, dx$.

Proof: Take $F(x) = \int_a^x f(x)\, dx$. The Mean Value Theorem for Derivatives (§3.2) says there is a value $c \in (a,b)$ where the tangent line for $F(x)$ is parallel to the secant line over the interval: $F'(c) = \frac{F(b)-F(a)}{b-a}$. By the First Fundamental Theorem, the left side is $F'(c) = f(c)$; and since $F(a) = 0$, the right side is $\frac{F(b)}{b-a} = \frac{1}{b-a}\int_a^b f(x)\, dx$, as desired.

In our example $f(x) = \sqrt{x}$, there is $c \in (0,4)$ with $f(c) = \sqrt{c} = f_{\text{ave}} = \frac{4}{3}$: i.e. $c = \frac{16}{9}$.

† This is not circular reasoning: rather, mathematics aims to turn physical intuitions into abstract definitions which we can reason with independent of intuition, checking whether our abstract objects behave as they intuitively should. If not, we need better definitions, which capture more of our intuition. If so, our non-intuitive mathematical predictions will probably hold in the physical world.

Substitution Method

Reversing the Chain Rule. As we have seen from the Second Fundamental Theorem (§4.3), the easiest way to evaluate an integral $\int_a^b f(x)\,dx$ is to find an antiderivative, the indefinite integral $\int f(x)\,dx = F(x) + C$, so that $\int_a^b f(x)\,dx = F(b)-F(a)$. Building on §3.9, we will find antiderivatives by reversing our methods of differentiation: here, we reverse the Chain Rule, $F(g(x))' = F'(g(x))\,g'(x)$.

For example, let us find the antiderivative:

$$\int x\cos(x^2)\,dx\,.$$

That is, for what function will the Derivative Rules produce $x\cos(x^2)$? We notice an inside function $g(x) = x^2$, and a factor x which is very close to the derivative $g'(x) = 2x$. In fact, we can get the exact derivative of the inside function if we multiply the factors by $\frac{1}{2}$ and 2:

$$x\cos(x^2) \;=\; \tfrac{1}{2}\cos(x^2)\cdot(2x)\,.$$

This is just the kind of derivative function produced by the Chain Rule:

$$F(g(x))' \;=\; F'(g(x))\cdot g'(x) \;=\; F'(x^2)\cdot(2x) \;\overset{??}{=}\; \tfrac{1}{2}\cos(x^2)\cdot(2x)\,.$$

We still need to find the outside function F. To remind us of the original inside function, we write $F(u)$, where the new variable u represents $u = g(x) = x^2$. We must get $F'(u) = \frac{1}{2}\cos(u)$, an easy antiderivative:

$$\int \tfrac{1}{2}\cos(u)\,du \;=\; F(u)+C \;=\; \tfrac{1}{2}\sin(u)+C\,.$$

Now we restore the original inside function to get our final answer:

$$\int \tfrac{1}{2}\cos(u)\,du \;=\; \tfrac{1}{2}\sin(u)+C \;=\; \tfrac{1}{2}\sin(x^2)+C\,.$$

The Chain Rule in Leibnitz notation (§2.5) reverses and checks the above computation. Writing $y = \frac{1}{2}\sin(u)$ and $u = x^2$:

$$\frac{dy}{dx} \;=\; \frac{dy}{du}\cdot\frac{du}{dx} \;=\; \frac{d}{du}\big(\tfrac{1}{2}\sin(u)\big)\cdot\frac{d}{dx}\big(x^2\big)$$
$$=\; \tfrac{1}{2}\cos(u)\cdot(2x) \;=\; \tfrac{1}{2}\cos(x^2)\cdot(2x) \;=\; x\cos(x^2)\,.$$

Notes by Peter Magyar `magyar@math.msu.edu`

Substitution Method

1. Given an antiderivative $\int h(x)\,dx$, try to find an inside function $g(x)$ such that $g'(x)$ is a factor of the integrand:
$$h(x) \;=\; f(g(x))\cdot g'(x).$$
This will often involve multiplying and dividing by a constant to get the exact derivative $g'(x)$. After factoring out $g'(x)$, sometimes the remaining factor needs to be manipulated to write it as a function of $u=g(x)$.

2. Using the symbolic notation $u=g(x)$, $\;du=\dfrac{du}{dx}\,dx=g'(x)\,dx$, write:
$$\int h(x)\,dx \;=\; \int f(g(x))\cdot g'(x)\,dx \;=\; \int f(u)\,du\,,$$
and find the antiderivative $\int f(u)\,du = F(u)+C$ by whatever method.

3. Restore the original inside function:
$$\int h(x)\,dx \;=\; \int f(u)\,du \;=\; F(u)+C \;=\; F(g(x))+C\,.$$

Examples

- $\int (3x{+}4)\sqrt{3x{+}4}\,dx$. The inside function is clearly $u=3x{+}4$, $du=3\,dx$, so:
$$\int (3x{+}4)\sqrt{3x{+}4}\,dx \;=\; \int \tfrac{1}{3}(3x{+}4)\sqrt{3x{+}4}\cdot 3\,dx$$
$$=\; \int \tfrac{1}{3}u\sqrt{u}\,du \;=\; \tfrac{1}{3}\int u^{3/2}\,du \;=\; \tfrac{1}{3}\,\tfrac{2}{5}u^{5/2}+C \;=\; \tfrac{2}{15}(3x{+}4)^{5/2}+C.$$

- $\displaystyle\int x\sqrt{3x{+}4}\,dx$. Again $u=3x{+}4$, so $\sqrt{3x{+}4}$ becomes $\sqrt{u}$, but we must still express the remaining factor x in terms of u. We solve $u=3x{+}4$ to obtain $x=\frac{1}{3}u-\frac{4}{3}$: that is, $x=\frac{1}{3}(3x{+}4)-\frac{4}{3}$:
$$\int x\sqrt{3x{+}4}\,dx \;=\; \int \tfrac{1}{3}\big(\tfrac{1}{3}(3x{+}4)-\tfrac{4}{3}\big)\sqrt{3x{+}4}\cdot 3\,dx \;=\; \int \tfrac{1}{3}\big(\tfrac{1}{3}u-\tfrac{4}{3}\big)\sqrt{u}\,du$$
$$=\; \int \tfrac{1}{9}u^{3/2}-\tfrac{4}{9}u^{1/2}\,du \;=\; \tfrac{1}{9}\,\tfrac{2}{5}u^{5/2}-\tfrac{4}{9}\,\tfrac{2}{3}u^{3/2}+C \;=\; \tfrac{2}{45}(3x{+}4)^{5/2}-\tfrac{8}{27}(3x{+}4)^{3/2}+C.$$

- $\displaystyle\int \frac{\sec^2(\sqrt{x})}{\sqrt{x}}\,dx\,.$ We take $u=\sqrt{x}=x^{1/2}$, $du=\frac{1}{2}x^{-1/2}\,dx=\frac{1}{2\sqrt{x}}\,dx$
$$\int \frac{\sec^2(\sqrt{x})}{\sqrt{x}}\,dx \;=\; \int 2\sec^2(\sqrt{x})\cdot\frac{1}{2\sqrt{x}}\,dx$$
$$=\; \int \sec^2(u)\,du \;=\; \tan(u)+C \;=\; \tan(\sqrt{x})+C.$$
Here we use the trig integrals from §3.9.

- $\displaystyle\int \frac{\sin(x)}{(1+\cos(x))^2}\,dx$. We cannot take the inside function $u=\sin(x)$, because its derivative $\cos(x)$ is not a factor of the integrand. We could take $u=\cos(x)$, but the best choice is $u=1+\cos(x)$, $du=-\sin(x)\,dx$:

$$\begin{aligned}\int \frac{\sin(x)}{(1+\cos(x))^2}\,dx &= -\int \frac{1}{(1+\cos(x))^2}\cdot(-\sin(x))\,dx\\ &= -\int \frac{1}{u^2}\,du \;=\; \frac{1}{u}+C \;=\; \frac{1}{1+\cos(x)}+C.\end{aligned}$$

- $\displaystyle\int \frac{1-\sqrt{x}}{\sqrt{1+\sqrt{x}}}\,dx$. Take $u=1+\sqrt{x}$, $du=\frac{1}{2\sqrt{x}}$, so $\sqrt{x}=u-1$, $1-\sqrt{x}=2-u$.

$$\begin{aligned}\int \frac{1-\sqrt{x}}{\sqrt{1+\sqrt{x}}}\,dx &= \int \frac{1-\sqrt{x}}{\sqrt{1+\sqrt{x}}}\,(2\sqrt{x})\cdot\frac{1}{2\sqrt{x}}\,dx\\ &= \int \frac{2-u}{\sqrt{u}}\,2(u-1)\,du \;=\; -2\int \frac{u^2-3u+2}{\sqrt{u}}\,du\\ &= -2\int u^{3/2}-3u^{1/2}+2u^{-1/2}\,du \;=\; -2(\tfrac{2}{5}u^{5/2}-2u^{3/2}+4u^{1/2})+C\\ &= -\tfrac{4}{5}(1+\sqrt{x})^{5/2}+4(1+\sqrt{x})^{3/2}-8(1+\sqrt{x})^{1/2}+C.\end{aligned}$$

Whew! Here we did not have the derivative factor $\frac{du}{dx}=\frac{1}{2\sqrt{x}}$ already present: we had to multipy and divide by it to get du, then express the remaining factors in terms of u. By luck, the resulting $\int f(u)\,du$ was do-able.

- $\displaystyle\int \sec^2(x)\tan(x)\,dx$. Here we could take $u=\tan(x)$, $du=\sec^2(x)\,dx$:

$$\begin{aligned}\int \sec^2(x)\tan(x)\,dx &= \int \tan(x)\cdot\sec^2(x)\,dx\\ &= \int u\,du \;=\; \tfrac{1}{2}u^2+C \;=\; \tfrac{1}{2}\tan^2(x)+C.\end{aligned}$$

Alternatively, use the inside function $z=\sec(x)$, $dz=\tan(x)\sec(x)\,dx$:

$$\begin{aligned}\int \sec^2(x)\tan(x)\,dx &= \int \sec(x)\cdot\tan(x)\sec(x)\,dx\\ &= \int z\,dz \;=\; \tfrac{1}{2}z^2+C \;=\; \tfrac{1}{2}\sec^2(x)+C.\end{aligned}$$

Thus $\frac{1}{2}\tan^2(x)$ and $\frac{1}{2}\sec^2(x)$ are two different antiderivatives, but what about the Antiderivative Uniqueness Theorem (§3.9)? In fact, the identity $\tan^2(x)+1=\sec^2(x)$ implies:

$$\tfrac{1}{2}\tan^2(x)+\tfrac{1}{2} \;=\; \tfrac{1}{2}\sec^2(x)\,.$$

These give the *same* antiderivative family: $\frac{1}{2}\tan^2(x)+C \;=\; \frac{1}{2}\sec^2(x)+C'$!

Substitution for definite integrals. We have, for $u = g(x)$:

$$\int_a^b f(g(x))\, g'(x)\, dx \;=\; \int_{g(a)}^{g(b)} f(u)\, du\,.$$

EXAMPLE: $\int_2^3 x(1{+}x^2)^5\, dx$. Taking $u = 1{+}x^2$, $du = 2x\, dx$:

$$\int_3^4 x(1{+}x^2)^5\, dx \;=\; \int_3^4 \tfrac{1}{2}(1{+}x^2)^5 \cdot 2x\, dx \;=\; \int_{1+3^2}^{1+4^2} \tfrac{1}{2}u^5\, du$$

$$=\; \tfrac{1}{12}\, u^6 \,\Big|_{u=10}^{u=17} \;=\; \tfrac{1}{12}10^6 - \tfrac{1}{12}17^6\,.$$

Integral Symmetry Theorem: If $f(x)$ is an odd function, meaning $f(-x) = -f(x)$, then $\int_{-a}^{a} f(x)\, dx = 0$.

Proof. By the Integral Splitting Rule (§4.2), we have:

$$\int_{-a}^{a} f(x)\, dx \;=\; \int_{-a}^{0} f(x)\, dx + \int_0^a f(x)\, dx\,.$$

Substituting $u = -x$, $du = (-1)\, dx$ in the first term, including in the limits of integration, and using $f(-x) = -f(x)$, we get:

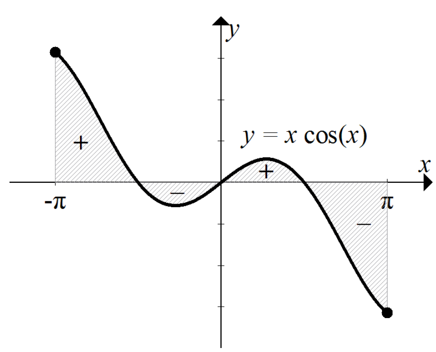
$$\int_{-a}^{0} f(x)\, dx \;=\; \int_{-a}^{0} -f(x)\cdot(-1)\, dx \;=\; \int_{-a}^{0} f(-x)\cdot(-1)\, dx$$

$$=\; \int_{-(-a)}^{-0} f(u)\, du \;=\; \int_a^0 f(u)\, du \;=\; -\int_0^a f(u)\, du \;=\; -\int_0^a f(x)\, dx\,.$$

The last equality holds because the variable of integration is merely suggestive, and can be changed arbitrarily. Therefore $\int_{-a}^{0} f(x)\, dx + \int_0^a f(x)\, dx = -\int_0^a f(x)\, dx + \int_0^a f(x)\, dx = 0$, as desired.

EXAMPLE: Evaluate the definite integral $\int_{-\pi}^{\pi} x\cos(x)\, dx$. Here substitution will not work, and it is difficult to find an antiderivative. But since $(-x)\cos(-x) = -(x\cos(x))$, the Theorem tells us the integral must be *zero*.

Geometrically, the integral is the signed area between the graph and the x-axis:



Since the function $f(x) = x\cos(x)$ is odd, the graph has rotational symmetry around the origin, and each negative area below the x-axis cancels a positive area above the x-axis.

Application: Heart Flow Rate. In a standard medical test to measure the rate of blood pumped by the heart, r liters/min, doctors inject a colored dye into a vein flowing toward the heart, then measure the concentration of dye in arterial blood as it is pumped out from the heart, $c(t)$ mg/liter after t minutes.

PROBLEM: Given the dye concentration function $c(t)$, determine the flow rate r.

Let the variable ℓ denote the volume of blood which has flowed through the artery since the start time. Assuming the (unknown) flow rate r is constant, we have $\ell = rt$. Let us define $C(\ell)$ to be the dye concentration after ℓ liters have flowed, so that $C(\ell) = C(rt) = c(t)$.

Now, the integral $\int_0^\infty C(\ell)\, d\ell$, summing up concentration (mg/liter) times increments of volume (liters), measures the total amount D of dye (mg):

$$D \;=\; \int_0^\infty C(\ell)\, d\ell.$$

Performing the substitution $\ell = rt$, $d\ell = r\, dt$, we have:

$$D \;=\; \int_0^\infty C(rt)\, r\, dt \;=\; r \int_0^\infty c(t)\, dt.$$

Then we may compute r as:

$$r \;=\; \frac{D}{\int_0^\infty c(t)\, dt}.$$

Since the total dye D is known (the amount injected), $c(t)$ is measured by the test, and $\int_0^\infty c(t)\, dt$ can be computed by Riemann sums,* we obtain flow rate r.

---

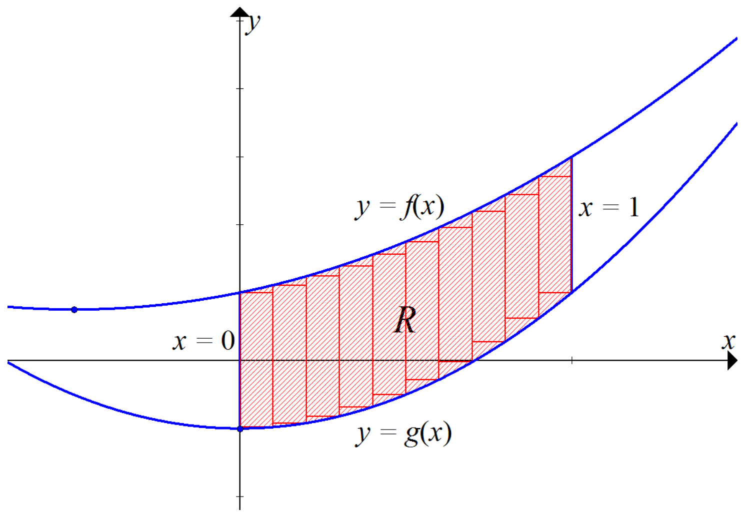
* Since $c(t) = 0$ after all the dye has passed, $\int_0^\infty c(t)\, dt$ can be cut off to a finite integral.

Region between two parabolas. We have seen that geometrically, the integral $\int_a^b f(x)\,dx$ computes the area between a curve $y=f(x)$ and an interval $x\in[a,b]$ on the x-axis (with area below the axis counted negatively). In Calculus II, we will show the versatility of the integral to compute all kinds of areas, lengths, volumes: almost any measure of size for a geometric object.

In this section, we compute more general areas: those between two given curves $y=f(x)$ and $y=g(x)$, usually with no boundary on the x-axis.

EXAMPLE: Consider the region with top boundary $y=f(x)=x^2+x+1$, bottom boundary $y=g(x)=2x^2-1$, left boundary the y-axis $x=0$, right boundary $x=1$:*

$$\begin{aligned} R &= \{\,(x,y) \text{ with } g(x)\le y\le f(x) \text{ and } x\in[0,1]\,\}. \\ &= \{\,(x,y) \text{ with } 2x^2-1\le y\le x^2+x+1 \text{ and } 0\le x\le 1\,\}. \end{aligned}$$



Here $y=g(x)=2x^2-1$ is a standard parabola shifted downward, with minimum point $x=0$. The curve $y=f(x)=x^2+x+1$ is roughly like its leading term $y=x^2$, a parabola opening upward; its minimum point satisfies $(x^2+x+1)'=2x+1=0$, i.e. $x=-\frac{1}{2}$.

To compute the area of R, we use the same geometric-numerical strategy as for the region under a single curve: split R into n thin vertical slices of width $\Delta x=\frac{1}{n}$, each approximately a rectangle; then add up the rectangle areas and take the limit as n becomes larger and larger. In the interval $x\in[0,1]$, we take sample points $x_1,\ldots,x_n$, one in each Δx increment. The slice at position x_i has height equal to the ceiling minus the floor, $f(x_i)-g(x_i)$, so:

$$\text{area of slice} \;\approx\; (\text{height})\times(\text{width}) \;=\; (f(x_i)-g(x_i))\,\Delta x,$$

Notes by Peter Magyar `magyar@math.msu.edu`

*We specify the region as the set of all points (x,y) which satisfy the given conditions.

and the total area is:

$$A_{[0,1]} \;=\; \lim_{n\to\infty}\sum_{i=1}^{n}(f(x_i)-g(x_i))\,\Delta x \;=\; \int_0^1 (f(x)-g(x))\,dx$$

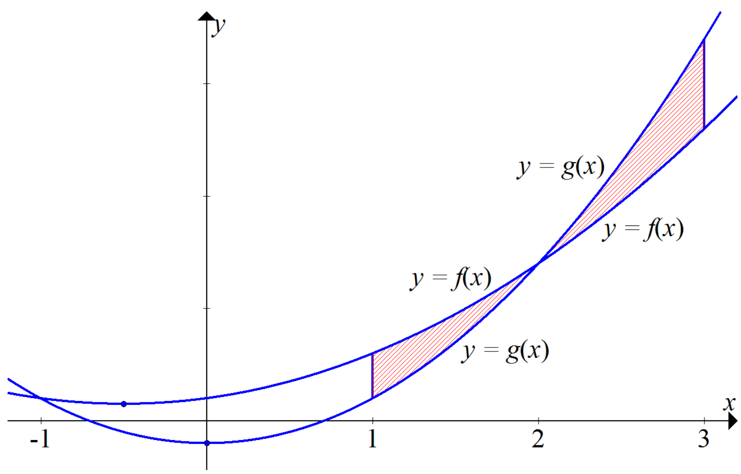
$$= \int_0^1 (x^2{+}x{+}1)-(2x^2{-}1)\,dx \;=\; \Big[2x{+}\tfrac12 x^2{-}\tfrac13 x^3\Big]_{x=0}^{x=1} \;=\; \tfrac{13}{6}\,.$$

EXAMPLE: Next, consider the region between the same curves $y=f(x)=x^2{+}x{+}1$ and $y=g(x)=2x^2{-}1$, but above the interval $x\in[1,3]$. To picture the region without a calculator, we determine the intersection points where the curves cross:

$$f(x)=g(x) \quad\Longleftrightarrow\quad x^2{+}x{+}1=2x^2{-}1 \quad\Longleftrightarrow$$

$$x^2{-}x{-}2=0 \quad\Longleftrightarrow\quad x=\tfrac{1\pm\sqrt{1^2-4(1)(-2)}}{2(1)}=\tfrac{1\pm3}{2}=-1 \text{ or } 2$$

by the Quadratic Formula. Only $x=2$ is relevant for our region above $x\in[1,3]$. At $x=1$ we have $g(1)<f(1)$, so to the left of $x=2$, our region is defined by $g(x)\le y\le f(x)$. At $x=3$, we have $f(3)<g(3)$, so to the right of $x=2$, it is $f(x)\le y\le g(x)$:



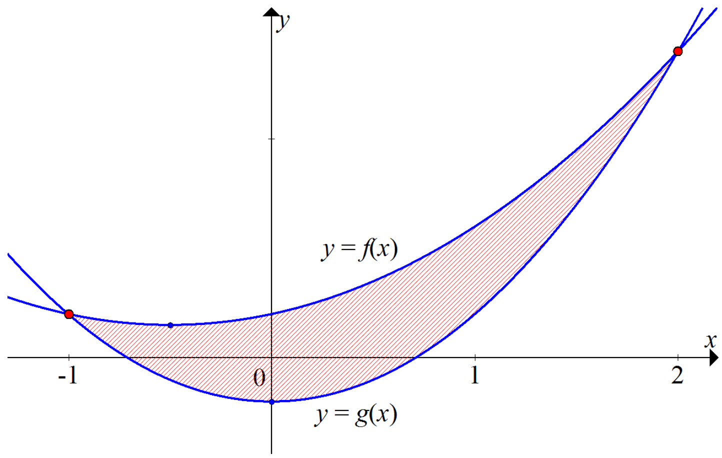
Repeating our previous area formula for the two parts of our region gives:

$$A_{[1,3]} \;=\; A_{[1,2]}+A_{[2,3]} \;=\; \int_1^2 (f(x)-g(x))\,dx+\int_2^3 (g(x)-f(x))\,dx$$

$$= \int_1^2 (2{+}x{-}x^2)\,dx \;+\; \int_2^3 (-2{-}x{+}x^2)\,dx \;=\; \tfrac76+\tfrac{11}{6} \;=\; 3.$$

EXAMPLE: Finally, we consider the same curves $y = f(x) = x^2{+}x{+}1$ and $y = g(x) = 2x^2{-}1$, but we take the entire finite region between them:

$$R' = \{(x, y) \text{ with } g(x) \le y \le f(x)\}.$$



Here the top boundary is $y = x^2 + x + 1$ and the bottom boundary is $y = 2x^2 - 1$, but we have not specified an x interval. However, we have already computed the intersection points $x = -1$ and $x = 2$, and the curves do not enclose any finite regions beyond these points. Thus:

$$A_{[-1,2]} \;=\; \int_{-1}^{2} (f(x) - g(x))\, dx = \tfrac{9}{2}\,.$$

We can generalize the above examples in:

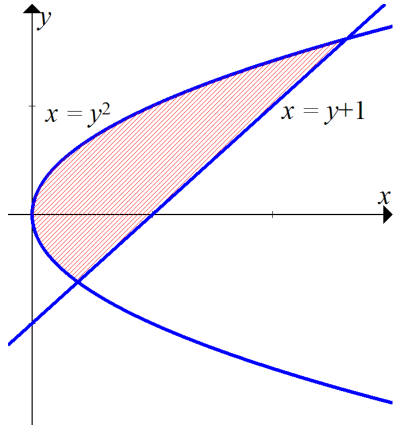
> *Theorem:* The area of the region enclosed between $f(x)$ and $g(x)$ for $x \in [a, b]$ is: $A \;=\; \displaystyle\int_a^b \big|f(x) - g(x)\big|\; dx\,.$

The absolute value signs ensure we take the integral of top minus bottom, regardless of which is which. In practice, we must find the intersection points where $f(x) = g(x)$, which split the integral into intervals where $g(x) \le f(x)$ versus $f(x) \le g(x)$.

Integrating with respect to y. Consider the region:

$$R \;=\; \{(x, y) \text{ with } y^2 \le x \le y{+}1\}\,.$$

Here the boudary curves are naturally graphs in which y is the independent variable: the right boundary is the line $x = f(y) = y{+}1$; and the left boundary is $x = g(y) = y^2$, a parabola opening to the right.



Understand: it is merely by habit that we consider y as a function of x. We can make x a function of y instead if it is more convenient, and the same formulas will work if we switch the roles of x and y. Thus, we find the intersection points: $y{+}1 = y^2$ when $y = \frac{1\pm\sqrt{5}}{2}$ by the Quadratic Formula. The area as:

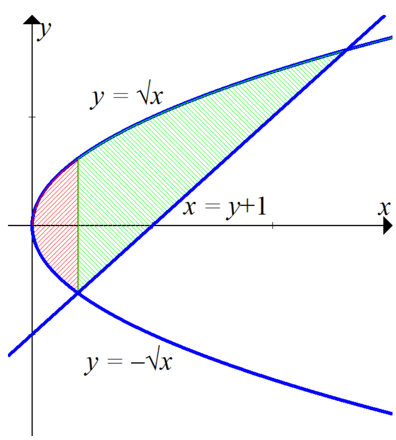
$$A \;=\; \int_{\frac{1-\sqrt{5}}{2}}^{\frac{1+\sqrt{5}}{2}} (y{+}1) - (y^2)\, dy \;=\; \left[\tfrac{1}{2}y^2 + y - \tfrac{1}{3}y^3\right]_{y=\frac{1-\sqrt{5}}{2}}^{y=\frac{1+\sqrt{5}}{2}} \;=\; \tfrac{5}{6}\sqrt{5}.$$

Here $((y{+}1) - (y^2))\, dy$ represents the area of the *horizontal* slice of the region at height y, with thickness dy.

To check this, we re-do it from our usual perspective, using x as the independent variable. This makes it more complicated, since we must consider the region as having three boundary graphs: upper boundary $y = \sqrt{x}$, lower right boundary $y = x{-}1$, and lower left boundary $y = -\sqrt{x}$. The intersection points are:

- Between $y = \sqrt{x}$ and $y = x{-}1$: $x = \frac{3+\sqrt{5}}{2}$ (upper right corner)
- Between $y = -\sqrt{x}$ and $y = x{-}1$: $x = \frac{3-\sqrt{5}}{2}$ (lower middle corner)
- Between $y = \sqrt{x}$ and $y = -\sqrt{x}$: $x = 0$ (left end)

These split the region into left and right parts:



The area is:

$$A \;=\; \int_0^{\frac{3-\sqrt{5}}{2}} (\sqrt{x}) - (-\sqrt{x})\, dx + \int_{\frac{3-\sqrt{5}}{2}}^{\frac{3+\sqrt{5}}{2}} (\sqrt{x}) - (x{-}1)\, dx\,,$$

which after much algebra gives the same answer as before: $\frac{5}{6}\sqrt{5}$.