Appendix I

Malcolm

Started 7 May 2024

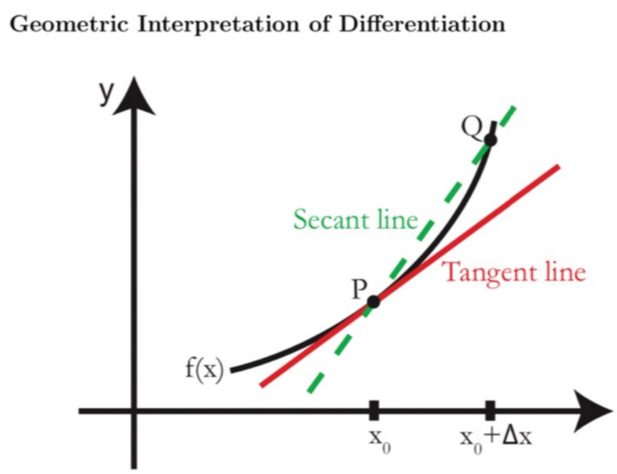
Contents

Appendix A

Single Variable Calculus/Calculus I and II

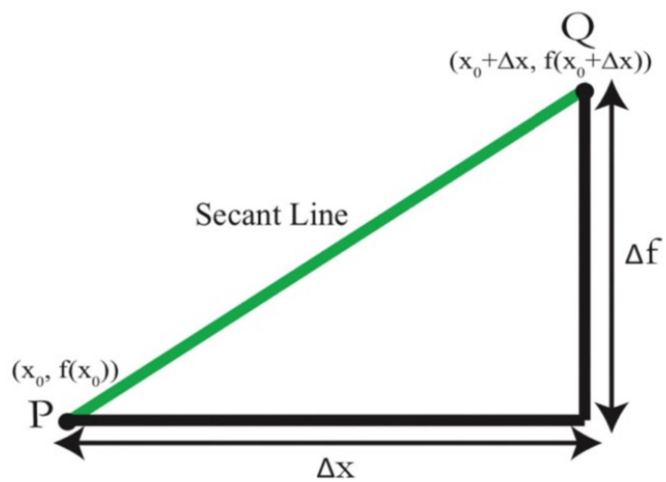
A.1 Differentiation

A.1.1 Definition of differentiation



The derivative of $f(x)$ at $x = x_0$ is the slope of the tangent line to $f(x)$ at the point $(x_0, f(x_0))$. Supposing PQ is a secant line of $f(x)$, the tangent line can be viewed as the limit of secant lines PQ as $Q \to P$ (P is fixed while Q varies).

Now consider a closer perspective of the secant line: Since the derivative is the



slope of the tangent line at P, it is also the slope of the secant line as $Q \to P$; therefore:

$$\underbrace{m}_{\text{derivative}} = \lim_{Q \to P} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}$$

Since we can further define $\Delta f = f(x_0 + \Delta x) - f(x_0)$, We can now formulate the derivative as:

$$m = f'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \underbrace{\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}}_{\textbf{Difference quotient}}$$

A machine could use this formula together with coordinates $(x_0.f(x_0))$ to draw a tangent line. This formula also serves as the basis for many other proofs for derivatives.

A.1.2 Derivative of $\frac{1}{x}$

As mentioned in A.1.1, the formula for the derivative:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

is fundamental to deriving many other functions; here we find the derivative of the function $f(x) = \frac{1}{x}$ at point $x = x_0$. First we formulate the gradient of the secant line:

$$\begin{aligned}
\frac{\Delta f}{\Delta x} &= \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \\
&= \frac{\frac{1}{x_0 + \Delta x} - \frac{1}{x_0}}{\Delta x} \\
&= \frac{(x_0 + \Delta x)(x_0)}{(x_0 + \Delta x)(x_0)} \frac{\frac{1}{x_0 + \Delta x} - \frac{1}{x_0}}{\Delta x} \\
&= \frac{x_0 - (x_0 + \Delta x)}{(x_0 + \Delta x)(x_0)(\Delta x)} \\
&= \frac{-\Delta x}{(x_0 + \Delta x)(x_0)(\Delta x)} \\
&= \frac{-1}{(x_0 + \Delta x)(x_0)}
\end{aligned}$$

Next, as Δx tends to zero:

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{-1}{(x_0 + \Delta x)(x_0)} = \boxed{-\frac{1}{x_0^2}}$$

A.1.3 Notation

Note the typical notation for the derivative comes from taking the limit as $\Delta x \to 0$:

$$\begin{aligned}
\frac{\Delta y}{\Delta x} &\to \frac{dy}{dx} && \text{(Leibniz' notation)} \\
\frac{\Delta f}{\Delta x} &\to f'(x_0) && \text{(Newton's notation)}
\end{aligned}$$

Note that:

$$\Delta y = \Delta f = f(x) - f(x_0) = f(x_0 + \Delta x) - f(x_0)$$

A.1.4 Derivative of x^n where $n = 1, 2, 3 \ldots$ (Power rule)

We plug $y = f(x)$ into the definiton of the difference quotient (see A.1.1):

$$\frac{\Delta y}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x} = \frac{(x + \Delta x)^n - x^n}{\Delta x}$$

Since $(x + \Delta x)^n$ can be written as:

$$(x + \Delta x)(x + \Delta x) \ldots (x + \Delta x) \quad \text{(n times)}$$

It can be rewritten as:

$$(x + \Delta x)^n = x^n + n(\Delta x)x^{n-1} + O(\Delta x)^2$$

where $O(\Delta x)^2$ represents all the terms containing $(\Delta x)^2, (\Delta x)^3$, and so on up till $(\Delta x)^n$. (Regarding $n(\Delta x)x^{n-1}$, consider that there were n different Δx's that one could choose to multiply by, so one gets this result n different ways.)

Returning to the difference quotient:

$$\frac{\Delta y}{\Delta x} = \frac{(x + \Delta x)^n - x^n}{\Delta x} = \frac{(x^n + n(\Delta x)x^{n-1} + O(\Delta x)^2) - x^n}{\Delta x} = nx^{n-1} + O(\Delta x)$$

(Where $O(\Delta x)$ represents all the terms containing $(\Delta x), (\Delta x)^2 \ldots (\Delta x)^n$.)

Taking the limit as $\Delta x \to 0$:

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = nx^{n-1}$$

and therefore,

$$\underbrace{\frac{d}{dx} x^n = nx^{n-1}}_{\textbf{power rule}}$$

A.1.5 Derivative of a Sum

Here we prove, where u and v are differentiable functions of x:

$$(u+v)'(x) = u'(x) + v'(x)$$

Note that $(u+v)(x) = u(x) + v(x)$

Proof: using the difference quotient:

$$\begin{aligned}
(u+v)'(x) &= \lim_{\Delta x \to 0} \frac{(u+v)(x+\Delta x) - (u+v)(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{u(x+\Delta x) + v(x+\Delta x) - u(x) - v(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \left\{ \frac{u(x+\Delta x) - u(x)}{\Delta x} + \frac{v(x+\Delta x) - v(x)}{\Delta x} \right\}
\end{aligned}$$

Because u and v are differentiable (and therefore continuous), the limit of the sum is the sum of the limits. Therefore:

$$\begin{aligned}
(u+v)'(x) &= \lim_{\Delta x \to 0} \frac{u(x+\Delta x) - u(x)}{\Delta x} + \lim_{\Delta x \to 0} \frac{v(x+\Delta x) - v(x)}{\Delta x} \\
&= u'(x) + v'(x)
\end{aligned}$$

A.1.6 Proof of limit for $\lim_{x\to 0} \frac{\sin x}{x}$

In order to compute specific formulas for the derivatives of $\sin x$ and $\cos x$, we run into the the limit $\lim_{\Delta x\to 0} \frac{\sin \Delta x}{\Delta x}$. Here we show that

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

Using θ as our parameter, consider:

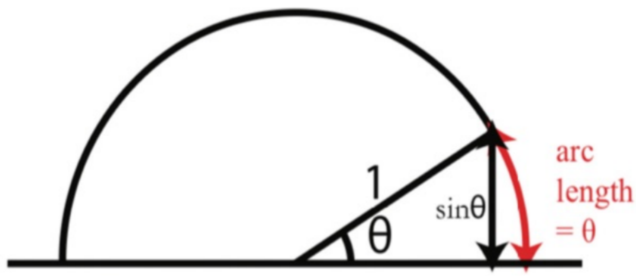


Figure: Radius 1, arc of angle θ

Notice that our function of interest $\frac{\sin\theta}{\theta}$ is the ratio of the edge length to the arc length. (Note the highlighted arc length is only equal to θ when θ is measured in **radians**)

Now consider what happens as θ becomes smaller.
When $\theta = \frac{\pi}{2}$ rad, $\sin\theta = 1$ and $\frac{\sin\theta}{\theta} = \frac{2}{\pi} \approx \frac{2}{3}$.
When $\theta = \frac{\pi}{4}$ rad, $\sin\theta = \frac{\sqrt{2}}{2}$ and $\frac{\sin\theta}{\theta} = \frac{2\sqrt{2}}{\pi} \approx \frac{9}{10}$.

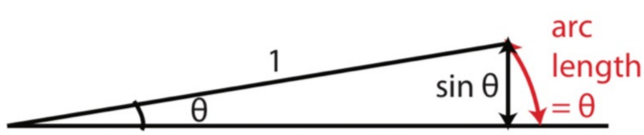


Figure: What happens as θ becomes very small?

As θ shrinks, the length of the segment $\sin\theta$ gets closer to the arc length θ ; thus we conclude:

$$\lim_{\theta\to 0} \frac{\sin\theta}{\theta} = 1 \quad \text{and}$$
$$\lim_{x\to 0} \frac{\sin x}{x} = 1$$

A.1.7 Proof of limit for $\lim_{x\to 0} \frac{1-\cos x}{x}$

In order to compute specific formulas for the derivatives of $\sin x$ and $\cos x$, we run into the the limit $\lim_{\Delta x\to 0} \frac{\cos \Delta x - 1}{\Delta x}$. Here we show that

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

Using θ as our parameter, consider:

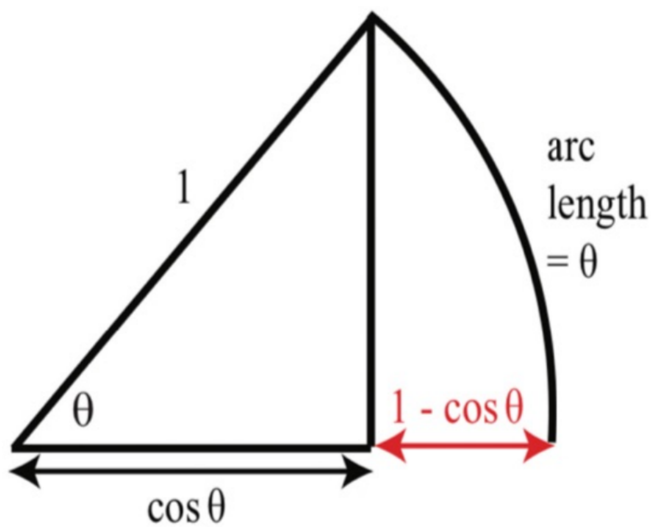


Figure: Radius 1, arc of angle θ

Notice that as $\theta \to 0$, the horizontal gap $1-\cos\theta$ gets much smaller compared to the length of the arc (This can be confirmed with a calculator or any graphing tool):

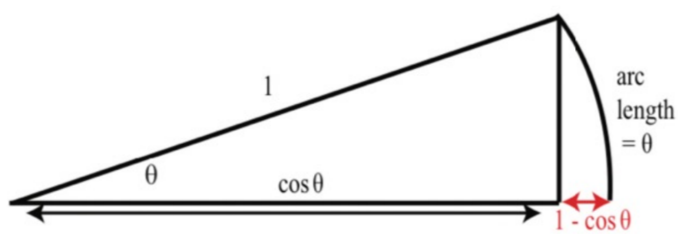


Figure: as θ becomes very small

(next page)

Since $|1 - \cos\theta|$ becomes increasingly smaller than θ as θ decreases, we can conclude:

$$\lim_{x \to 0} \frac{1 - \cos\theta}{\theta} = \lim_{x \to 0} \frac{\cos\theta - 1}{\theta} = 0 \quad \text{and}$$
$$\lim_{x \to 0} \frac{1 - \cos x}{x} = \lim_{x \to 0} \frac{\cos x - 1}{x} = 0$$

A.1.8 Derivative of $\sin x$, Algebraic proof; Angle formulas

Here we compute a specific formula for the derivative of $\sin x$. We begin with the definition of the derivative/difference quotient:

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

using the angle formula (proof below):

$$\sin(a + b) = \sin(a)\cos(b) + \sin(b)\cos(a)$$

we get:

$$\begin{aligned}
\frac{d}{dx}\sin x &= \lim_{\Delta x \to 0} \frac{(\sin x \cos \Delta x + \sin \Delta x \cos x) - \sin x}{\Delta x} \\
&= \lim_{\Delta x \to 0} \left[\frac{\sin x(\cos \Delta x - 1)}{\Delta x} + \frac{\sin \Delta x \cos x}{\Delta x} \right] \\
&= \lim_{\Delta x \to 0} \sin x \left(\frac{\cos \Delta x - 1}{\Delta x} \right) + \lim_{\Delta x \to 0} \cos x \left(\frac{\sin \Delta x}{\Delta x} \right) \\
&= \sin x \lim_{\Delta x \to 0} \left(\frac{\cos \Delta x - 1}{\Delta x} \right) + \cos x \lim_{\Delta x \to 0} \left(\frac{\sin \Delta x}{\Delta x} \right)
\end{aligned}$$

Using the fact that

$$\begin{aligned}
&\lim_{\Delta x \to 0} \frac{\cos \Delta x - 1}{\Delta x} = 0 \quad \text{and} \\
&\lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x} = 1
\end{aligned}$$

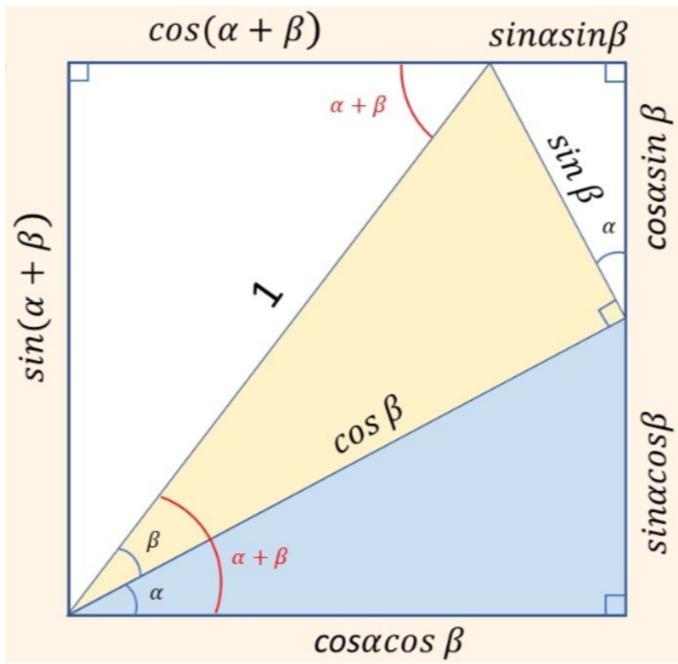
We conclude:

$$\begin{aligned}
\frac{d}{dx}\sin x &= \sin x \lim_{\Delta x \to 0} \left(\frac{\cos \Delta x - 1}{\Delta x} \right) + \cos x \lim_{\Delta x \to 0} \left(\frac{\sin \Delta x}{\Delta x} \right) \\
&= \cos x
\end{aligned}$$

(Proof of angle formula on next page)

Proof of:

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \sin\beta\cos\alpha \quad \text{and}$$
$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$



A.1.9 Derivative of $\cos x$, Algebraic proof

Here we compute a specific formula for the derivative of $\cos x$. We begin with the definition of the derivative/difference quotient:

$$\frac{d}{dx}\cos x = \lim_{\Delta x \to 0} \frac{\cos(x + \Delta x) - \cos x}{\Delta x}$$

using the angle formula

$$\cos(a + b) = \cos a \cos b - \sin a \sin b$$

we get:

$$\begin{aligned}
\frac{d}{dx}\cos x &= \lim_{\Delta x \to 0} \frac{\cos(x + \Delta x) - \cos x}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{(\cos x \cos \Delta x - \sin x \sin \Delta x) - \cos x}{\Delta x} \\
&= \lim_{\Delta x \to 0} \left[\cos x \frac{\cos \Delta x - 1}{\Delta x} - \sin x \frac{\sin \Delta x}{\Delta x}\right] \\
&= \cos x \lim_{\Delta x \to 0} \frac{\cos \Delta x - 1}{\Delta x} - \sin x \lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x}
\end{aligned}$$

Using the fact that

$$\begin{aligned}
&\lim_{\Delta x \to 0} \frac{\cos \Delta x - 1}{\Delta x} = 0 \quad \text{and} \\
&\lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x} = 1
\end{aligned}$$

We conclude:

$$\begin{aligned}
\frac{d}{dx}\cos x &= \cos x \lim_{\Delta x \to 0} \frac{\cos \Delta x - 1}{\Delta x} - \sin x \lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x} \\
&= -\sin x
\end{aligned}$$

A.1.10 Derivative of $\sin x$, Geometric proof

This appendix also contains an algebraic proof for the derivative of $\sin x$; however while that proof was valid, it did not make use of the definition of the sine function. Here we prove that the derivative of $\sin\theta$ is $\cos\theta$ directly from the definition of the sine function as the ratio $\frac{|\text{opposite}|}{|\text{hypotenuse}|}$ of the a right triangle. Consider a circle:

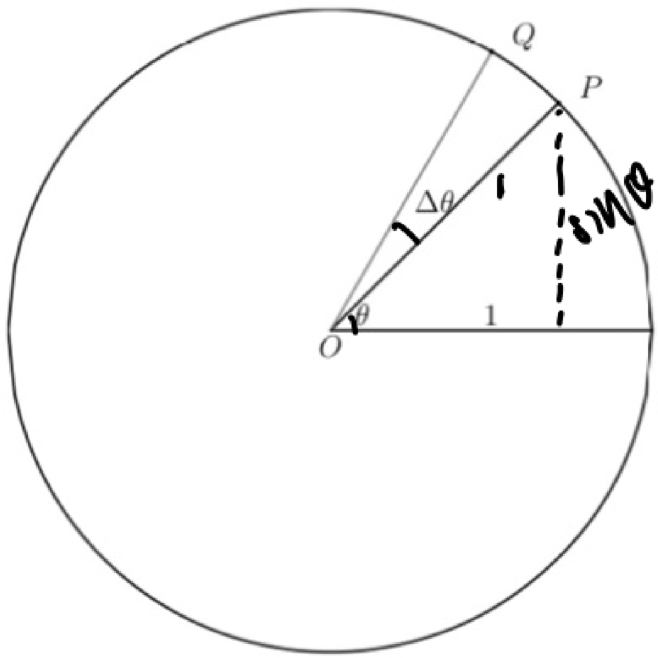


Figure: point $P = \sin\theta$

Notice that $\sin\theta$ is the vertical distance between P and the x axis, we increment θ slightly by the addition of $\Delta\theta$. Notice now that point Q is the point on the unit circle at angle $\theta + \Delta\theta$, and that the y-coordinate of Q is $\sin(\theta + \Delta\theta)$. To find the rate of change of $\sin\theta$ with respect to θ we just need to find the rate of change of $y = \sin\theta$.

(next page)

Now consider a close-up view of segment PQ:

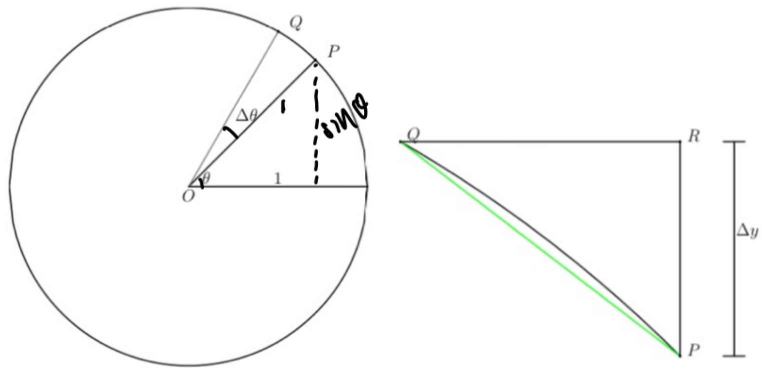


Figure: When $\Delta\theta$ is small, **segment** $PQ \approx$ **arc** PQ

Notice that $\Delta y = |PR|$ and **segment** PQ is a straight line that approximates to **arc** PQ as $\Delta\theta$ becomes smaller. Also note that since the circle is of radius 1, arc $PQ = \Delta\theta$ and segment $PQ \approx \Delta\theta$ (as $\Delta\theta$ decreases).

Since $\Delta\theta$ is small, segment PQ is (nearly) tangent to the circle, and angle $\angle OPQ$ is (nearly) a right angle. We can say therefore that $\angle RPQ$ and θ are (nearly) congruent angles.

The arc $PQ = \Delta\theta$ approximates to $|PQ|$(hypotenuse), and $\angle RPQ$ approximates to θ (As $\Delta\theta$ becomes small). Now consider that $\cos\theta \approx \cos(\angle RPQ)$ and therefore

$$\cos\theta \approx \frac{|PR|}{\Delta\theta} = \frac{\Delta y}{\Delta\theta} = \frac{\sin(\theta+\Delta\theta)-\sin\theta}{\Delta\theta} \quad \text{(As } \Delta\theta \text{ becomes small)}$$

As $\Delta\theta \to 0$ **arc** PQ approximates **segment** PQ more closely and $\angle OPQ$ approximates a right angle more closely. This means that $\frac{\sin(\theta+\Delta\theta)-\sin\theta}{\Delta\theta}$ more closely approximates $\cos\theta$. We can therefore conclude that:

$$\underbrace{\lim_{\Delta\theta\to 0}\frac{\sin(\theta+\Delta\theta)-\sin\theta}{\Delta\theta}}_{\frac{d}{dx}\sin\theta} = \cos\theta$$

A.1.11 Product rule

Here we prove the product rule, the derivative of the product of two functions:

$$(uv)' = u'v + uv'$$

We start with the definition of the derivative:

$$\begin{aligned}
\frac{d}{dx}(uv) &= \lim_{\Delta x \to 0} \frac{(uv)(x + \Delta x) - (uv)(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{u(x + \Delta x)v(x + \Delta x) - u(x)v(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{u(x + \Delta x)v(x + \Delta x) - u(x)v(x) + u(x + \Delta x)v(x) - u(x + \Delta x)v(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{u(x + \Delta x)(v(x + \Delta x) - v(x)) + v(x)(u(x + \Delta x) - u(x))}{\Delta x} \\
&= \lim_{\Delta x \to 0} \left[u(x + \Delta x) \frac{v(x + \Delta x) - v(x)}{\Delta x} \right] + v(x) \left[\lim_{\Delta x \to 0} \frac{u(x + \Delta x) - u(x))}{\Delta x} \right] \\
&= u(x)v'(x) + v(x)u'(x) \\
&= u'(x)v(x) + u(x)v'(x)
\end{aligned}$$

Note we used the fact that:

$$\lim_{\Delta x \to 0} u(x + \Delta x) = u(x)$$

This works because we assume u is differentiable and continuous.

A.1.12 Quotient rule

Here we derive a specific function for differentiating quotients (fractions), the quotient rule:

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

(Note that this can simply be seen as an extension of taking the derivative of (uv^{-1}), where the product rule can also be used.)

We start with the definition of the derivative:

$$\left(\frac{u}{v}\right)' = \lim_{\Delta x \to 0} \frac{\frac{u(x+\Delta x)}{v(x+\Delta x)} - \frac{u(x)}{v(x)}}{\Delta x}$$

First we simplify the numerator using:

$$\begin{aligned} u(x+\Delta x) - u(x) &= \Delta u \quad (\text{not } u(\Delta x)) \\ u(x+\Delta x) &= u(x) + \Delta u \end{aligned}$$

where

$$\begin{aligned} \frac{u(x+\Delta x)}{v(x+\Delta x)} - \frac{u(x)}{v(x)} &= \frac{u+\Delta u}{v+\Delta v} - \frac{u}{v} \\ &= \frac{v(u+\Delta u) - u(v+\Delta v)}{v(v+\Delta v)} \\ &= \frac{(\Delta u)v - u(\Delta v)}{v(v+\Delta v)} \end{aligned}$$

Now back to the derivative:

$$\begin{aligned} \lim_{\Delta x \to 0} \frac{\frac{u(x+\Delta x)}{v(x+\Delta x)} - \frac{u(x)}{v(x)}}{\Delta x} &= \lim_{\Delta x \to 0} \frac{\frac{(\Delta u)v - u(\Delta v)}{v(v+\Delta v)}}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{1}{\Delta x} \frac{(\Delta u)v - u(\Delta v)}{v(v+\Delta v)} \\ &= \lim_{\Delta x \to 0} \frac{\left(\frac{\Delta u}{\Delta x}\right) v - u\left(\frac{\Delta v}{\Delta x}\right)}{v(v+\Delta v)} \\ &= \frac{u'v - uv'}{v^2} \end{aligned}$$

A.1.13 Quotient rule derived from Product and Chain rule

As mentioned earlier, the Quotient rule is simply an extension of the product rule and the chain rule:

$$\left(\frac{f(x)}{g(x)}\right)' = (f(x)(g(x))^{-1})'$$

First we find the derivative of $(g(x))^{-1}$ using the chain rule:

$$\begin{aligned}\frac{d}{dx}(g(x))^{-1} &= \frac{d}{dg}(g(x))^{-1} \cdot \frac{d}{dx}g(x) \\ &= -(g(x))^{-2} \cdot g'(x)\end{aligned}$$

Therefore:

$$\begin{aligned}\left(\frac{f(x)}{g(x)}\right)' &= (f(x)(g(x))^{-1})' \\ &= f'(x)(g(x))^{-1} + f(x)((g(x))^{-1})' \\ &= f'(x)(g(x))^{-1} + f(x)(-(g(x))^{-2}g'(x)) \\ &= [f'(x)g(x) - f(x)g'(x)]\,(g(x))^{-2} \\ &= \underbrace{\frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}}_{\text{product rule}}\end{aligned}$$

A.1.14 Chain rule

The composition or "chain" rule tells us how to find the derivative of a composition of functions like $f(g(x))$ (note this is **not** the same as $fg(x)$).

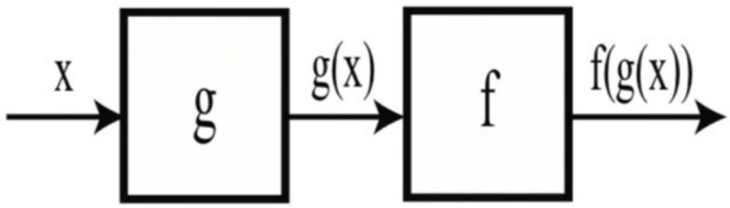


Figure: Composition of functions: $(f \circ g)(x) = f(g(x))$

The chain rule can be viewed as using an intermediate variable to find a derivative. Consider trying to find $\frac{dy}{dt}$:

$$\frac{\Delta y}{\Delta t} = \frac{\Delta y}{\Delta x} \cdot \frac{\Delta x}{\Delta t}$$
$$\lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} = \underbrace{\frac{dy}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt}}_{\textbf{chain rule}}$$

The chain rule can also be written as

$$\frac{d}{dx}(f \circ g)(x) = \frac{d}{dx} f(g(x)) = \underbrace{f'(g(x)) \cdot g'(x)}_{\frac{dy}{dx} \cdot \frac{dx}{dt}}$$

A.1.15 Chain rule example: $\frac{d}{dt} \sin^n t$

Example: Consider $y = \sin^n t$ (in the form of $y(t)$), where we want to find $\frac{dy}{dt}$; we can introduce a function x where $x = \sin t$, and now $y = x^n$ (notice it is now in the form $y(x(t))$. Now applying the chain rule:

$$\begin{aligned}
\frac{dy}{dt} = \frac{d}{dt} \sin^n t &= \frac{dy}{dx} \cdot \frac{dx}{dt} \\
&= \frac{dy}{dx} x^n \cdot \frac{d}{dt} \sin t \\
&= n x^{n-1} \cdot \cos t \\
&= n \sin^{n-1} t \cdot \cos t
\end{aligned}$$

A.1.16 Chain rule example: $\frac{d}{dt}\sin(nt)$

Example: Consider $y = \sin(nt)$ (in the form of $y(t)$), where we want to find $\frac{dy}{dt}$; we can introduce a function x where $x = nt$, and now $y = \sin x$ (notice it is now in the form $y(x(t))$. Now applying the chain rule:

$$\begin{aligned}\frac{dy}{dt} = \frac{d}{dt}\sin nt &= \frac{dy}{dx}\cdot\frac{dx}{dt} \\ &= \frac{dy}{dx}\sin x\cdot\frac{d}{dt}nt \\ &= \cos x\cdot n \\ &= n\cos(nt)\end{aligned}$$

A.1.17 Derivative of $y = x^a$, where a is rational: Rational exponent rule (via implicit differentiation)

Here we prove the rational exponent rule, where

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

Where a is a rational number (i.e $\frac{m}{n}$)

Consider $y = x^{\frac{m}{n}}$, where m and n are integers; we want to compute $\frac{dy}{dx}$. Unlike most proofs, starting with the definition of derivative does not work here (since we get stuck on $(x + \Delta x)^{\frac{m}{n}}$). We get around this by:

$$\begin{aligned} y &= x^{\frac{m}{n}} \\ y^n &= x^{\frac{m}{n} \cdot n} \\ y^n &= x^m \end{aligned}$$

Now we perform **implicit differentiation**, where we take the derivative of a function that cannot be explicitly expressed as $f(x)$, in this case y^n:

$$\begin{aligned} y^n &= x^m \\ \frac{d}{dx}y^n &= \frac{d}{dx}x^m \end{aligned}$$

using the chain rule to simplify $\frac{d}{dx}y^n$:

$$\begin{aligned} \frac{d}{dx}y^n &= \frac{d}{dy}y^n \cdot \frac{dy}{dx} \\ &= ny^{n-1}\frac{dy}{dx} \end{aligned}$$

we now have:

$$\begin{aligned} \frac{d}{dx}y^n &= \frac{d}{dx}x^m \\ ny^{n-1}\frac{dy}{dx} &= mx^{m-1} \end{aligned}$$

(next page)

Now we can rearrange:

$$\begin{aligned} ny^{n-1}\frac{dy}{dx} &= mx^{m-1} \\ \frac{dy}{dx} &= \frac{mx^{m-1}}{ny^{n-1}} \end{aligned}$$

Finally we can replace y with $x^{\frac{m}{n}}$:

$$\begin{aligned} \frac{dy}{dx} &= \frac{mx^{m-1}}{ny^{n-1}} \\ &= \frac{m}{n}\left(\frac{x^{m-1}}{x^{\frac{m}{n}\cdot(n-1)}}\right) \\ &= \frac{m}{n}x^{(m-1)-\frac{m(n-1)}{n}} \\ &= \frac{m}{n}x^{\frac{n(m-1)}{n}-\frac{m(n-1)}{n}} \\ &= \frac{m}{n}x^{\frac{m-n}{n}} \\ &= \frac{m}{n}x^{(\frac{m}{n}-1)} \end{aligned}$$

For any rational number a, the derivative of x^a is ax^{a-1}.

A.1.18 Derivative of the inverse of a function

Consider $f(x) = \sqrt{x}$ plotted against its inverse function:

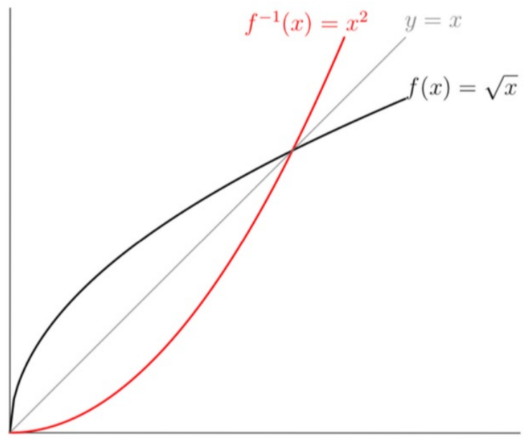


Figure: $f^{-1}(x)$ is a reflection of $f(x)$ across the line $y = x$

Since an inverse function f^{-1} can generally be found by exchanging the $x-$ and $y-$coordinates of f (a reflection across $y = x$), one might posit that if $\frac{dy}{dx}$ is the slope of a line tangent to f, then

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

would be the slope of a line tangent to f^{-1}. Here we prove this through implicit differentiation; consider:

$$y = f(x)$$
$$f^{-1}(y) = x$$
$$\frac{d}{dx}(f^{-1}(y)) = \frac{d}{dx}x = 1$$

by the chain rule:

$$\frac{d}{dy}(f^{-1}(y))\frac{dy}{dx} = 1$$
$$\frac{d}{dy}(f^{-1}(y)) = \frac{1}{\frac{dy}{dx}}$$

A.1.19 Derivative of $\tan(x)$ and $\arctan(x)/\tan^{-1}(x)$

Derivative of $\tan(x)$**:**
First we prove the derivative of $\tan(x)$ using the product rule:

$$\begin{aligned}\tan(x) &= \frac{\sin(x)}{\cos(x)}\\ \frac{d}{dx}\tan(x) &= \frac{d}{dx}\frac{\sin(x)}{\cos(x)}\end{aligned}$$

Now by product rule:

$$\begin{aligned}\frac{d}{dx}\frac{\sin(x)}{\cos(x)} &= \frac{\cos(x)\cdot\cos(x) - \sin(x)\cdot(-\sin(x))}{(\cos(x))^2}\\ &= \frac{\sin^2(x)+\cos^2(x)}{\cos^2(x)}\\ &= \frac{1}{\cos^2(x)} = \sec^2(x)\end{aligned}$$

Derivative of $\arctan(x)/\tan^{-1}(x)$: Here we use implicit differentiation to find the derivative of $y = \tan^{-1}(x) = \arctan(x)$:

$$\begin{aligned}y &= \tan^{-1}(x)\\ \tan(y) &= x\\ \frac{d}{dx}\tan(y) &= \frac{d}{dx}x\\ \frac{d}{dy}\tan(y)\cdot\frac{dy}{dx} &= 1 \quad \text{(Chain rule)}\\ \frac{1}{\cos^2(y)}\cdot\frac{dy}{dx} &= 1\\ \frac{dy}{dx} &= \cos^2(y)\end{aligned}$$

Now we can substitute in $y = \tan^{-1}(x)$ to get

$$\frac{dy}{dx} = \cos^2(\tan^{-1}(x))$$

Although this result is valid, it can still be simplified greatly.
(next page)

Our earlier result was:

$$\frac{dy}{dx} = \cos^2(y)$$
$$\frac{dy}{dx} = \cos^2(\tan^{-1}(x))$$

We can use geometry to simplify this result; consider:

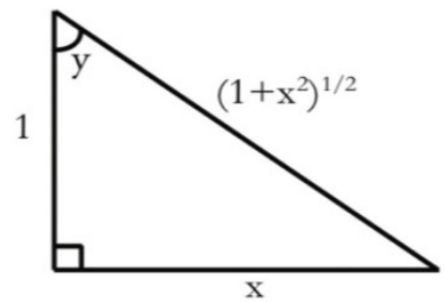


Figure: Triangle where $\tan(y) = x$

Using the Pythagorean theorem we know that the hypotenuse of this theoretical triangle is

$$h = \sqrt{1+x^2}$$

so now we can compute

$$\cos(y) = \frac{1}{\sqrt{1+x^2}}$$
$$\cos^2(y) = \frac{1}{1+x^2}$$

Therefore,

$$\frac{d}{dx}\tan^{-1}(x) = \frac{d}{dx}\arctan(x) = \frac{dy}{dx} = \cos^2(y) = \frac{1}{1+x^2}$$

A.1.20 Derivative of $\arcsin(x)/\sin^{-1}(x)$

We can solve for the derivative of $y = \arcsin(x) = \sin^{-1}(x)$ by implicit differentiation:

$$
\begin{aligned}
y &= \sin^{-1}(x) \\
\sin(y) &= x \\
\frac{d}{dx}\sin(y) &= \frac{d}{dx}x \\
\frac{d}{dy}\sin(y)\cdot\frac{dy}{dx} &= 1 \\
(\cos(y))\cdot\frac{dy}{dx} &= 1 \\
\frac{dy}{dx} &= \frac{1}{\cos(y)}
\end{aligned}
$$

Since

$$
\begin{aligned}
\sin^2(y) + \cos^2(y) &= 1 \\
\cos(y) &= \sqrt{1 - \sin^2(y)}
\end{aligned}
$$

(Notice that we made a choice between a positive and negative square root when solving for $\cos(y)$. We chose the positive square root as $\sin^{-1}(x)$ is usually defined to have outputs between $-\pi/2$ and $\pi/2$, a range in which the cosine function is always positive.)

We can now simplify the derivative:

$$
\begin{aligned}
\frac{dy}{dx} &= \frac{1}{\cos(y)} \\
&= \frac{1}{\sqrt{1 - \sin^2(y)}} \\
&= \frac{1}{\sqrt{1 - x^2}}
\end{aligned}
$$

A.1.21 Intuition and definition of e (Derivative of a^x part 1)

Here we attempt to provide an introduction and a sense of intuition to Euler's number e. This is done by attempting to find the derivative of a^x, where e is a natural result that falls out of the math during the derivation.

We begin with the goal of calculating the derivative $\frac{d}{dx}a^x$. We start with the definition of the derivative:

$$\begin{aligned}
\frac{d}{dx}a^x &= \lim_{\Delta x \to 0} \frac{a^{x+\Delta x} - a^x}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{a^x a^{\Delta x} - a^x}{\Delta x} \\
&= \lim_{\Delta x \to 0} a^x \frac{a^{\Delta x} - 1}{\Delta x} \\
&= a^x \lim_{\Delta x \to 0} \frac{a^{\Delta x} - 1}{\Delta x}
\end{aligned}$$

We can see that $\frac{d}{dx}a^x$ is a^x multiplied by some value that we don't yet know. We denote that value by $M(a)$:

$$M(a) = \lim_{\Delta x \to 0} \frac{a^{\Delta x} - 1}{\Delta x}$$

So

$$\frac{d}{dx}a^x = M(a)a^x$$

Notice that when we plug $x = 0$ into the definition of the derivative $\frac{d}{dx}a^x$:

$$\begin{aligned}
\left.\frac{d}{dx}a^x\right|_{x=0} &= \left.\lim_{\Delta x \to 0} \frac{a^{x+\Delta x} - a^x}{\Delta x}\right|_{x=0} \\
&= \lim_{\Delta x \to 0} \frac{a^{0+\Delta x} - a^0}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{a^{\Delta x} - 1}{\Delta x} = M(a)
\end{aligned}$$

(we could also observe that $\left.\frac{d}{dx}a^x\right|_{x=0} = M(a)a^0 = M(a)$)

Notice that $M(a)$ is the value of the derivative of a^x when $x = 0$.
(next page)

On the necessity for e: The result

$$\left.\frac{d}{dx}a^x\right|_{x=0} = M(a) = \lim_{\Delta x \to 0} \frac{a^{\Delta x} - 1}{\Delta x}$$

indicates that $M(a)$ can be thought of as the gradient of a^x at $x = 0$, meaning we only need to know the value of the gradient at $x = 0$ to know the gradient at any point on the curve (note that the shape of the curve depends on the value of a, resulting in different tangent lines and different $M(a)$):

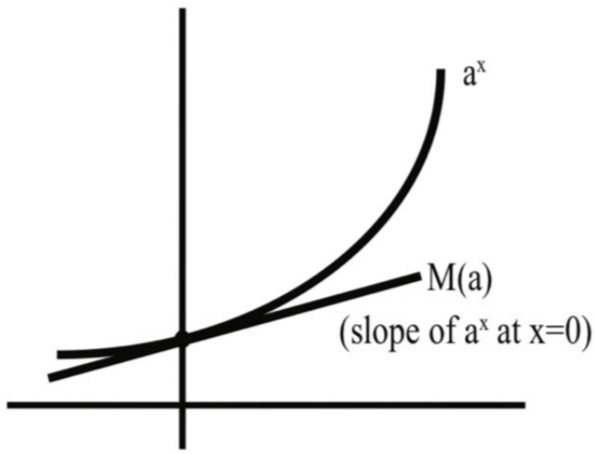


Figure: Geometric definition of $M(x)$

However, one runs into difficulty when attempting to exactly identify $M(a)$. This issue is circumvented by the introduction of e.

Definition of e:
We need to know what $M(a)$ is to find the derivative of a^x. It turns out the easiest way to understand $M(a)$ is to give up trying to calculate it and to *define* a number e as the number where $M(e) = 1$. This would lead to the result

$$\frac{d}{dx}e^x = e^x$$

The gradient of the tangent line to $y = e^x$ at $x = 0$ is 1. This can be confirmed:

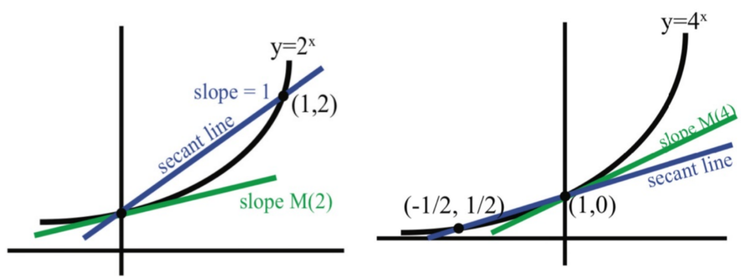
$$\left.\frac{d}{dx}e^x\right|_{x=0} = e^0 = 1$$

(next page)

On the existence and identity of e:
We define e to be the number where:

$$\frac{d}{dx}e^x = e^x, \quad M(e) = 1$$

Now we elaborate on why e exists. We know that the $M(a)$ (gradient of a^x at $x = 0$) increases as a increases. Consider:



Left: $M(2)$ compared with secant line of gradient 1
Right: $M(4)$ compared with secant line of gradient 1

The figures plot $y = 2^x$ (left) and $y = 4^x$ (right) against $y = x + 1$ (line of gradient 1 passing through (1,0)). Notice that $M(2) < 1$ while $M(4) > 1$. Using this we can conclude that e exists, and that $2 < e < 4$.

In conclusion, we define e as the unique number where:

$$M(e) = 1$$

or

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

or

$$\frac{d}{dx}e^x = 1 \quad \text{at } x = 0$$

A.1.22 Differentiating $e^{f(x)}$

$$\begin{aligned}\frac{d}{dx}e^{f(x)} &= \frac{d}{df}e^{f(x)} \cdot \frac{df}{dx} \quad \text{(chain rule)}\\ &= e^{f(x)} \cdot \frac{df}{dx}\\ &= f'(x)e^{f(x)}\end{aligned}$$

A.1.23 Natural Logarithm and its derivative

Intuition and definition:
Recall that:

$$M(a) = \lim_{\Delta x \to 0} \frac{a^{\Delta x} - 1}{\Delta x}$$

Is the value for the derivative $\frac{d}{dx}a^x = M(a)a^x$, and is equivalent to the value of the derivative of a^x at $x = 0$ ($\frac{d}{dx}a^x\big|_{x=0}$ gives $M(a)$). Also recall how this naturally leads to e ($M(e) = 0$ so $\frac{d}{dx}e^x = e^x$) (see A.1.21).

Here we introduce and derive the natural logarithm. The natural log function $\ln(x)$ is the inverse of the function e^x; it is defined as follows:

$$\text{If} \quad y = e^x, \quad \text{then} \quad \ln(y) = x$$

or

$$\text{If} \quad w = \ln(x), \quad \text{then} \quad x = e^w$$

Properties: The natural log has additive properties:

$$\ln(x_1 x_2) = \ln(x_1) + \ln(x_2)$$

This can be proved with simple concepts, consider:

$$x_1 = a \quad \text{and} \quad x_2 = b$$

so, since $e^{\ln(x)}$ is equivalent to $f(f^{-1}(x)) = x$:

$$x_1 = e^{\ln(a)} \quad \text{and} \quad x_2 = e^{\ln(b)}$$

therefore

$$\begin{aligned} \ln(x_1 x_2) &= \ln(e^{\ln(a)} e^{\ln(b)}) \\ &= \ln(e^{(\ln(a) + \ln(b))}) \\ &= (\ln(a) + \ln(b)) = \ln(x_1) + \ln(x_2) \end{aligned}$$

Other properties also include:

$$\ln(1) = 0 \quad (\ln(e^0) = 0)$$

$$\ln(e) = 1 \quad (\ln(e^1) = e)$$

Also note that the nature of $\ln(x)$ as the inverse of e^x means that it lies entirely to the right of the y-axis, and has a gradient of 1 at $x = 0$ ($\frac{dx}{dy}e^x\Big|_{x=0}$). Next we find the derivative of $\ln(x)$.
(next page)

Derivative of natural log: We use implicit differentiation:

$$\begin{aligned}
y &= \ln(x) \\
e^y &= x \\
\frac{d}{dx}e^y &= \frac{d}{dx}x \\
\frac{d}{dy}e^y \cdot \frac{dy}{dx} &= 1 \quad \text{(chain rule)} \\
e^y \cdot \frac{dy}{dx} &= 1 \\
\frac{dy}{dx} &= \frac{1}{e^y}
\end{aligned}$$

Since $y = \ln(x)$ and $e^y = x$:

$$\begin{aligned}
\frac{dy}{dx} = \frac{d}{dx}\ln(x) &= \frac{1}{e^y} = \frac{1}{x} \\
\frac{d}{dx}\ln(x) &= \frac{1}{x}
\end{aligned}$$

And for function $f(x)$, differentiating $\ln(f(x))$:

$$\begin{aligned}
\frac{d}{dx}\ln(f(x)) &= \frac{d}{df}\ln(f(x)) \cdot \frac{df}{dx} \quad \text{(chain rule)} \\
&= \frac{1}{f(x)} \cdot f'(x) \\
&= \frac{f'(x)}{f(x)}
\end{aligned}$$

A.1.24 Proof of $\ln(a^x) = x\ln(a)$

Here we prove:

$$\ln(a^x) = x\ln(a)$$

Since $e^{\ln(x)} = x$,

$$\begin{aligned}
\ln(a^x) &= \ln((e^{\ln(a)})^x) \\
&= \ln(e^{x\ln(a)}) \\
&= x\ln(a)
\end{aligned}$$

A.1.25 Derivative of a^x part 2

Here we outline the the derivative of a^x, this can be done through different methods:

Method 1: Conversion to base e

Since (by chain rule) $\frac{d}{dx}e^{f(x)} = f'(x)e^{f(x)}$, we convert the base to e:

$$\begin{aligned}
\frac{d}{dx}a^x &= \frac{d}{dx}(e^{\ln(a)})^x \\
&= \frac{d}{dx}e^{\ln(a)x} \\
&= \frac{d}{dx}(\ln(a)x) \cdot e^{\ln(a)x} \\
&= \ln(a)a^x
\end{aligned}$$

Recall that $\frac{d}{dx}a^x = M(a) \cdot a^x$. We can now define $M(a)$:

$$M(a) = \ln(a)$$

Method 2: Logarithmic

We can use $\frac{d}{dx}\ln(f(x)) = \frac{f'(x)}{f(x)}$ to avoid changing bases and differentiate implicitly:

$$\begin{aligned}
y &= a^x \\
\ln(y) &= \ln(a^x) \\
\frac{d}{dx}\ln(y) &= \frac{d}{dx}(x\ln(a)) \\
\frac{d}{dy}\ln(y)\frac{dy}{dx} &= \ln(a) \quad \text{(chain rule)} \\
\frac{1}{y}\frac{dy}{dx} &= \ln(a) \\
\frac{dy}{dx} &= \ln(a)y \\
&= \ln(a)a^x
\end{aligned}$$

Note therefore that both methods yield the same solution:

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

A.1.26 $\frac{d}{dx}x^r = rx^{r-1}$ for any *real* r (another power rule proof)

Earlier in this appendix are sections detailing proofs of the power rule regarding the derivative of x^r for integer and rational values of r. Here we show two proofs that the power rule

$$\frac{d}{dx}x^r = rx^{r-1}$$

applies for any *real* value of r.

Method 1: Base e
Since $x = e^{\ln(x)}$,

$$\begin{aligned}
x^r &= (e^{\ln(x)})^r \\
\frac{d}{dx}x^r &= \frac{d}{dx}(e^{r\ln(x)}) \\
&= \frac{d}{dx}(r\ln(x))e^{r\ln(x)} \\
&= \frac{r}{x}e^{r\ln(x)} \\
&= \frac{r}{x}x^r \\
&= rx^{r-1}
\end{aligned}$$

Method 2: Logarithmic
We start with the fact that, defining $f(x) = x^r$

$$\begin{aligned}
(\ln(f))' &= \frac{f'}{f} \\
f' &= f(\ln(f))'
\end{aligned}$$

Now we find $(\ln(f))'$:

$$\begin{aligned}
(\ln(f))' &= (\ln(x^r))' \\
&= (r\ln(x))' \\
&= \frac{r}{x}
\end{aligned}$$

Therefore,

$$f' = f(\ln(f))' = (x^r)\frac{r}{x} = rx^{r-1}$$

A.1.27 On the value of e (Moving exponent)

Here we present a method for computing the value of e. This comes from attempting to find the value of the limit

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

Note that this limit contains a moving exponent n. To begin evaluating the limit we first apply a logarithm to turn the exponent into a multiple.

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

Now consider the idea that

$$\lim_{n\to\infty}\frac{1}{n} = \lim_{\Delta x\to 0}\Delta x$$

We apply that idea here, changing the limit to substitute $\frac{1}{n}$ for Δx:

$$\begin{aligned}
\lim_{n\to\infty}\left[n\ln\left(1+\frac{1}{n}\right)\right] &= \lim_{\Delta x\to 0}\left[\frac{1}{\Delta x}\ln\left(1+\Delta x\right)\right] \\
&= \lim_{\Delta x\to 0}\left[\frac{\ln(1+\Delta x)-\ln(1)}{\Delta x}\right] \qquad (\text{since } \ln(1)=\ln(e^0)=0) \\
&= \frac{d}{dx}\ln(x)\bigg|_{x=1} = \frac{1}{x}\bigg|_{x=1} = 1
\end{aligned}$$

Now we return to the original problem:

$$\begin{aligned}
\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n &= \lim_{n\to\infty}\left(e^{\ln(1+\frac{1}{n})}\right)^n \\
&= \lim_{n\to\infty} e^{n\ln(1+\frac{1}{n})} \\
&= e^{\lim_{n\to\infty}\left[n\ln(1+\frac{1}{n})\right]} \\
&= e^1 = e
\end{aligned}$$

Therefore, since e happens to be the result where

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

We can use this to compute a value of e. For instance

$$\left(1+\frac{1}{10000}\right)^{10000} \approx 2.7182$$

A.1.28 Derivatives of Hyperbolic Sine and Cosine

Hyperbolic sine:

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

Hyperbolic cosine:

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

Intuition for these hyperbolic functions comes from Euler's notation, see [TBD].

The derivative of $\sinh(x)$ is $\cosh(x)$ and vice versa:

$$\frac{d}{dx}\sinh(x) = \frac{d}{dx}\left(\frac{e^x - e^{-x}}{2}\right) = \left(\frac{e^x + e^{-x})}{2}\right) = \cosh(x)$$

$$\text{likewise,} \quad \frac{d}{dx}\cosh(x) = \sinh(x)$$

Also note an important identity:

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

Proof:

$$\begin{aligned}
\cosh^2(x) - \sinh^2(x) &= \left(\frac{e^x + e^{-x}}{2}\right)^2 - \left(\frac{e^x - e^{-x}}{2}\right)^2 \\
&= \frac{1}{4}(e^{2x} + 2e^x e^{-x} + e^{-2x}) - \frac{1}{4}(e^{2x} - 2e^x e^{-x} + e^{-2x}) \\
&= 1
\end{aligned}$$

Some intuition:
Letting $u = \cosh(x)$ and $v = \sinh(x)$, then

$$u^2 - v^2 = 1$$

Which is the equation of a hyperbola

A.2 Applications of differentiation

A.2.1 Intuition for Linear approximation

Graphical intuition
Consider a curve $f(x)$; now consider an x-coordinate x_0. One can intuit that the curve is approximately the same as its tangent line $y = f(x_0) + f'(x_0)(x - x_0)$:

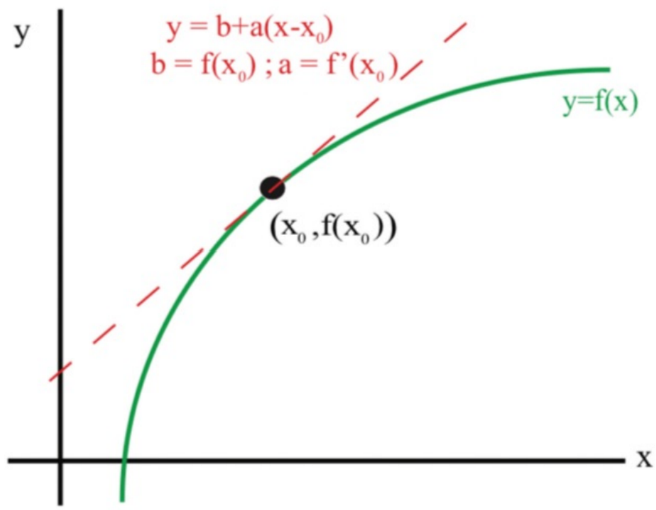


Figure: approximation grows less accurate moving away from x_0

The tangent line approximates $f(x)$ near the tangent point x_0. However, moving away from x_0, the approximation

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

grows less accurate.
(next page)

Intuition from the derivative
Another way to understand the formula for linear approximation involves the definition of the derivative:

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}$$

This can be interpreted to mean:

$$\frac{\Delta f}{\Delta x} \approx f'(x_0) \quad \text{as } \Delta x \to 0$$

Note that Δx refers to $x - x_0$, allowing us to rewrite:

$$\begin{aligned}
\frac{f(x) - f(x_0)}{x - x_0} &\approx f'(x) \\
f(x) - f(x_0) &\approx f'(x_0)(x - x_0) \\
f(x) &\approx f(x_0) + f'(x_0)(x - x_0)
\end{aligned}$$

A.2.2 Linear approximations near 0 for Sine, Cosine, and Exponential functions

Here we describe the linear approximation of several common functions at 0. To simplify things we use base point $x_0 = 0$ and assume that $x \approx 0$; this gives us a simplified general formula:

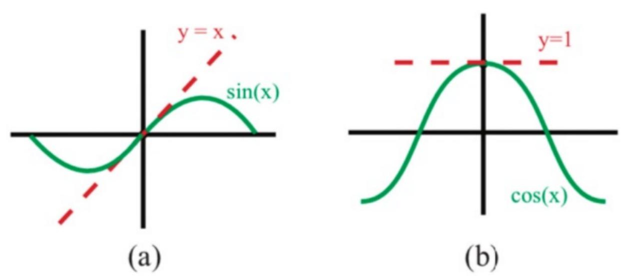
$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$
$$\text{becomes}$$
$$f(x) \approx f(0) + f'(0)(x)$$

Note therefore that these approximations won't work when x is not near 0.
We summarise the values of $f'(x), f(x), f'(0)$ and $f(0)$ as follows:

$f(x)$	$f'(x)$	$f(0)$	$f'(0)$
$\sin(x)$	$\cos(x)$	0	1
$\cos(x)$	$-\sin(x)$	1	0
e^x	e^x	1	1

Plugging the above values into our simplified formula $f(x) \approx f(0) + f'(0)(x)$, we get:

1. $\sin(x) \approx x$ (if $x \approx 0$) (see figure (a))
2. $\cos(x) \approx 1$ (if $x \approx 0$) (see figure (b))
3. $e^x \approx 1 + x$ (if $x \approx 0$)



Figures: Graphical intuition for linear approximations when $x \approx 0$

A.2.3 Linear approximations near 0 for $\ln(1+x)$ and $(1-x)^r$

Here we describe the linear approximation of $\ln(1+x)$ and $(1-x)^r$ at 0. To simplify things we use base point $x_0 = 0$ and assume that $x \approx 0$; this gives us a simplified general formula:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$
becomes
$$f(x) \approx f(0) + f'(0)(x)$$

We summarise the values of $f'(x), f(x), f'(0)$ and $f(0)$ as follows:

$f(x)$	$f'(x)$	$f(0)$	$f'(0)$
$\ln(1+x)$	$\dfrac{1}{1+x}$	0	1
$(1+x)^r$	$r(1-x)^{r-1}$	1	r

This gives us the approximations:

1. $\ln(1+x) \approx x \quad (\text{if } x \approx 0)$
2. $(1-x)^r \approx 1 + rx \quad (\text{if } x \approx 0)$

Note that we approximate $\ln(1+x)$ because $\ln(x) \to -\infty$ as $x \to 0$. Similarly, we approximate $(1+x)^r$ because apparently for some values of r x^r is not well behaved when $x = 0$. If we really need an approximation for x^r we can get one via change of variables.

Example of change of variables

Approximating $\ln(x)$ at $x \approx 1$ via the linear approximation formula:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

Where $x_0 = 1$ in this case, gives us $\ln(x) \approx x - 1$. Notice that by substituing $u \approx 1 + x$ into the result it yields:

$$\ln(x) \approx x - 1 \quad \text{for } x \approx 1$$
becomes
$$\underbrace{\ln(1+u) \approx u}_{\text{derived above}} \quad \text{for } u \approx 0$$

Therefore, knowing $\ln(x) \approx x - 1$ for $x \approx 1$, one can approximate $\ln(1+u)$ by change of variables. (vice versa works too)

A.2.4 Quadratic Approximation - Definition and Intuition

Formula
Quadratic approximation can be seen as an extension of linear approximation. The formula for quadratic approximation is as follows:

$$f(x) \approx \underbrace{f(x_0) + f'(x_0)(x - x_0)}_{\text{Linear part}} + \underbrace{\frac{f''(x_0)}{2}(x - x_0)^2}_{\text{Quadratic part}} \quad \text{for } (x \approx x_0)$$

Intuition
Consider attempting to approximate a function that is a parabola. Intuitively, if the graph of a function is a parabola, that function *is a quadratic function* and would therefore be best approximated by an identical quadratic function. Consider:

$$f(x) = a + bx + cx^2; \quad f'(x) = b + 2cx; \quad f''(x) = 2c$$

Notice that by plugging in base point x_0, we can find formulas for the constants of the original function:

$$\begin{aligned} f(x_0) = a + bx_0 + cx_0^2 \quad &\implies \quad a = f(x_0) - bx_0 - cx_0^2 \\ f'(x_0) = b + 2cx_0 \quad &\implies \quad b = f'(x_0) - 2cx_0 \\ f''(x_0) = 2c \quad &\implies \quad c = \frac{f''(x_0)}{2} \end{aligned}$$

Now we attempt to find $f(x)$ (assuming $x = x_0$):

$$\begin{aligned} f(x) &= a + bx + cx^2 \\ &= f(x_0) - bx_0 - cx_0^2 + bx + cx^2 \quad \text{(plug in } a\text{)} \\ &= f(x_0) + b(x - x_0) + c(x^2 - x_0^2) \\ &= f(x_0) + (f'(x_0) - 2cx_0)(x - x_0) + c(x^2 - x_0^2) \quad (b \text{ now}) \\ &= f(x_0) + f'(x_0)(x - x_0) - 2cx_0x + 2cx_0^2 + c(x^2 - x_0^2) \\ &= f(x_0) + f'(x_0)(x - x_0) + c(x_0^2 - 2x_0x + x^2) \\ &= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 \quad (c \text{ now}) \end{aligned}$$

Note that quadratic approximation would therefore obviously perfectly approximate a *quadratic* function. The point of quadratic approximation is to approximate an otherwise oddly shaped function with a simpler polynomial.

One can think of quadratic approximation as an attempt to *fit* a parabola to a specific point on a function. Just as linear approximation attempts to fit a straight line to a point on a function.

Geometric significance As mentioned, the quadratic approximation gives the best-fit parabola to a function. Here we consider $\cos(x)$:

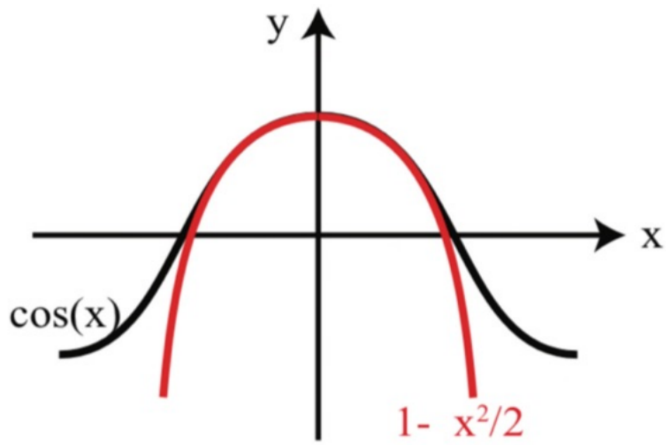


Figure: Quadratic approximation to $\cos(x)$

Consider a *linear* approximation to $\cos(x)$ near $x_0 = 0$, which would approximate the graph by a straight horizontal line $y = 1$, which obviously isn't a good approximation.

Now consider a similar *quadratic* approximation, as illustrated. The geometric intuition of both linear and quadratic approximation are similar, but quadratic approximation evidently allows for better approximation in this case.

A.2.5 Quadratic approsimations near 0

As per the formula for Quadratic Approximation:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0) \quad \text{where } x \approx x_0$$

Here we list the quadratic approximations at $x_0 = 0$ for a few common functions; we compute the second derivatives for the following functions, listed below:

$f(x)$	$f'(x)$	$f''(x)$	$f(0)$	$f'(0)$	$f''(0)$
$\sin(x)$	$\cos(x)$	$-\sin(x)$	0	1	0
$\cos(x)$	$-\sin(x)$	$-\cos(x)$	1	0	-1
e^x	e^x	e^x	1	1	1
$\ln(1+x)$	$\frac{1}{1+x}$	$-\frac{1}{(1+x)^2}$	0	1	-1
$(1+x)^r$	$r(1+x)^{r-1}$	$r(r-1)(1+x)^{r-1}$	1	r	$r(r-1)$

Plugging in the values of $f(0), f'(0)$, and $f''(0)$:

1. $\sin(x) \approx x \quad (\text{if } x \approx 0)$
2. $\cos(x) \approx a - \frac{x^2}{2} \quad (\text{if } x \approx 0)$
3. $e^x \approx 1 + x + \frac{1}{2}x^2 \quad (\text{if } x \approx 0)$
4. $\ln(1+x) \approx x - \frac{1}{2}x^2 \quad (\text{if } x \approx 0)$
5. $(1+x)^r \approx 1 + rx + \frac{r(r-1)}{2}x^2 \quad (\text{if } x \approx 0)$

A.2.6 Degree n approximation

Recall Linear and Quadratic approximation:

1. Linear: $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$
2. Quadratic: $f(x) \approx f(x_0) + f'(x_0)(x - x_0) + f''(x_0)(x - x_0)$

When $x \approx x_0$, x_0 being a predefined x-coordinate.

Linear and Quadratic approximations describe 1 and 2 degree approximations (degree referring to the highest derivative). Here we define higher degree approximations, with intuition similar to that of Linear and Quadratic approximation.

Here we focus on finding approximations *near the value* $x_0 = 0$; the calculations for the general case are very similar. Consider a general equation for a higher-order polynomial (a general formula for the approximation):

$$A(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_{n-1} x^{n-1} a_n x^n$$

Consider using derivatives, and the fact that $x_0 = 0$ to solve for coefficients a:

$$\begin{aligned}
A(x_0) = A(0) &= a_0 + a_1(0) + a_2(0) + \ldots + a_n(0) & \\
&= a_0 + 0 + 0 + \ldots + 0 & a_0 &= A(0) \\
A'(0) &= a_1 + 2a_2(0) + 3a_3(0) + \ldots + (n-1)a_n(0) & a_1 &= A'(0) \\
A^{(2)}(0) &= 2a_2 + 6a_3(0) + \ldots + (n-1)(n-2)a_n(0) & a_2 &= \frac{A^{(2)}(0)}{2} \\
A^{(3)}(0) &= 6a_3 + \ldots + (n)(n-1)(n-1)a_n(0) & a_3 &= \frac{A^{(3)}(0)}{6} \\
&\vdots & &\vdots \\
A^{(n)}(0) &= [(n)(n-1)(n-2) \cdot \ldots \cdot (2)(1)]a_n & a_n &= \frac{A^{(n)}(0)}{n!}
\end{aligned}$$

So our approximation would look something like (for $x \approx x_0 = 0$):

$$f(x) \approx A(x_0) = A(0) + A'(0)x + \frac{A^{(2)}(0)}{2} x^2 + \frac{A^{(3)}(0)}{6} x^3 + \ldots + \frac{A^{(n)}(0)}{n!} x^n$$

(next page)

Notice that the derivative of $a_n x^n$ is multiplied by a constant decreasing in increments of 1:

$$\begin{aligned}
\frac{d}{dx} a_n x^n &= (n) a_n x^{n-1} \\
\frac{d^2}{dx^2} a_n x^n &= (n)(n-1) a_n x^{n-2} \\
&\vdots \\
\frac{d^n}{dx^n} a_n x^n &= (n)(n-1)\ldots(3)(2)(1) a_n \\
&= n! a_n
\end{aligned}$$

Together, we have a general formula for each term:

$$a_i \approx \frac{f^{(i)}}{i!}$$

With this we can conclude (when $x \approx x_0 = 0$):

$$f(x) \approx f(0) + f'(0)x + \frac{f^{(2)}(0)}{2} x^2 + \ldots + \frac{f^{(n-1)}(0)}{(n-1)!} x^{n-1} + \frac{f^{(n)}(0)}{n!} x^n$$

A.2.7 Mean Value Theorem (MVT)

Here we introduce the Mean Value Theorem (MVT);

$$\frac{f(b) - f(a)}{b - a} = f'(c) \quad \text{(for some } c, a < c < b\text{)}$$

Provided that f is **differentiable** on $a < x < b$, and **continuous** on $a \leq x \leq b$.

Geometric inutition of MVT: Consider a graph of $f(x)$:

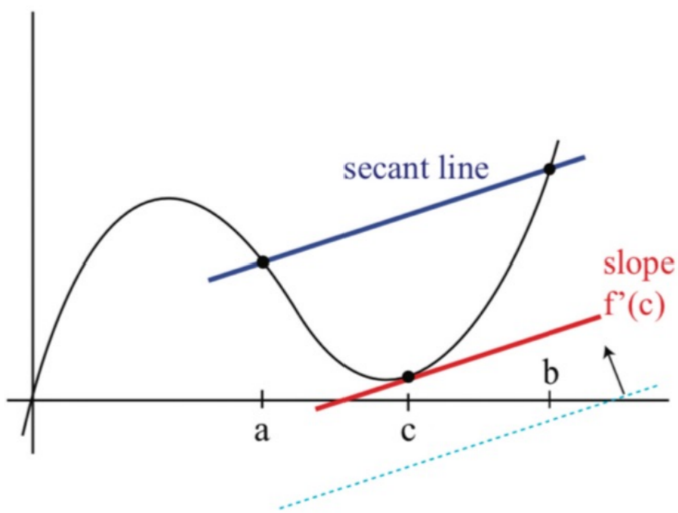


Figure: A point c exists where the $m_{\text{tangent}} = m_{\text{secant}}$

The MVT essentially states that there is some point c between a and b where the gradient of the secant line joining points $(a, f(a))$ and $(b, f(b))$ is equal to that of the slope of the tangent at that point (only if the graph is continuous and differentiable between points a and b).

Geometrically this can be seen by shifting a line parallel to the secant line vertically and noticing that it eventually touches a point between a and b.
(next page)

MVT only works if a function is continuous and differentiable:
Note that a discontinuity in the function (not continuous) or in the gradient of the function (not differentiable) would prevent the MVT from holding. Consider graphing $y = |x|$:

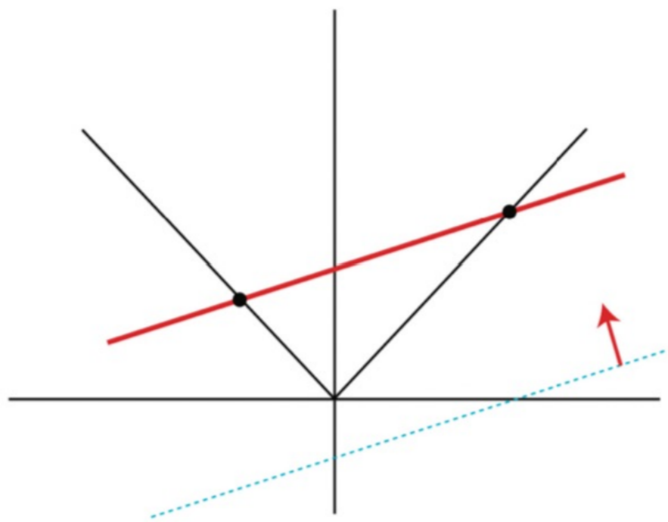


Figure: $y = |x|$ is not differentiable between a and b, MVT doesn't hold

Inuitively, there isn't a gradual change in gradient, meaning that the 'range' of $f'(x)$ between a and b doesn't include all values between $f'(a)$ and $f'(b)$.

A.2.8 Taylor's theorem from MVT

Here we use the Mean Value Theorem (MVT) to derive Taylor's theorem. First recall the MVT, which states that given a continuous function f on a closed interval $[a,b]$ which is differentiable on (a,b), there is a number c in (a,b) such that

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

Rearranging terms, we can make this equation look similar to a that of a linear approximation for $f(b)$ using a as a base point.

$$\begin{aligned} \text{from MVT:} \quad & f(b) = f(a) + f'(c)(b - a) \\ \text{linear approximation:} \quad & f(b) \approx f(a) + f'(a)(b - a) \end{aligned}$$

Notice that the only difference between a linear approximation is that the term $f'(a)$ has been *replaced* by $f'(c)$ for some point c in order to achieve an **exact** equality. However remember that the MVT only gives the existence of such a point c, and not a method to find c.

We interpret this as meaning that the difference between $f(b)$ and $f(a)$ is given by an expression resembling the *next* term in the Taylor polynomial. Here $f(a)$ is a "0-th degree" Taylor polynomial. A n-th order Taylor polynomial $P(x)$ for $f(x)$ with base point $x = a$ has the form:

$$P(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2 + \ldots + \frac{f^{(n)}(a)}{n!}(x - a)^n$$

Taylor's theorem for a first-degree Taylor polynomial is similarly:

$$f(b) = \underbrace{[f(a) + f'(a)(b - a)]}_{\text{1st degree taylor polynomial}} + \frac{f''(c)}{2}(b - a)^2$$

Generalising to an n-th degree Taylor polynomial:

n-th degree approximation:

$$f(b) \approx P(b) = f(a) + f'(a)(b - a) + \frac{f''(a)}{2}(b - a)^2 + \ldots + \frac{f^{(n)}(a)}{n!}(b - a)^n$$

Using Taylor's theorem:

$$\begin{aligned} f(b) = f(a) + f'(a)(b - a) + \frac{f''(a)}{2}(b - a)^2 + \ldots \\ \ldots + \frac{f^{(n)}(a)}{n!}(b - a)^n + \frac{f^{(n+1)}(a)}{(n+1)!}(b - a)^{n+1} \end{aligned}$$

Taylor's theorem, therefore, is the idea that given n-th degree Taylor polynomial:

$$f(b) - P(b) = \frac{f^{(n+1)}(a)}{(n+1)!}(b - a)^{n+1}$$

A.2.9 Difference between linear approximation and MVT

Here we elaborate on the differences between the Mean Value Theorem (MVT) and linear approximation for the sake of intuition.

A linear approximation can be expressed as (for some base point a):

$$\begin{aligned} f(x) &\approx f(a) + f'(a)(x-a) \\ f(x) - f(a) &\approx f'(a)\Delta x \\ \frac{\Delta f}{\Delta x} &\approx f'(a) \end{aligned}$$

The MVT can be expressed similarly:

$$\begin{aligned} \frac{f(b)-f(a)}{b-a} &= f'(c) \quad \text{for some } c \text{ where } a < c < b \\ f(b) &= f(a) + f'(c)(b-a) \\ \frac{f(b)-f(a)}{b-a} &= f'(c) \end{aligned}$$

The MVT tells us that $\frac{\Delta f}{\Delta x}$ is exactly equal to $f'(c)$ for some c between a and b. This means that the 'average change' on the interval $[a, b]$ is between the minimum and maximum values of $f'(x)$ (This works because f is continuous and differentiable on the interval):

$$\min_{a \leq x \leq b} \leq \frac{f(b)-f(a)}{b-a} = f'(c) \leq \max_{a \leq x \leq b}$$

Illustrated:

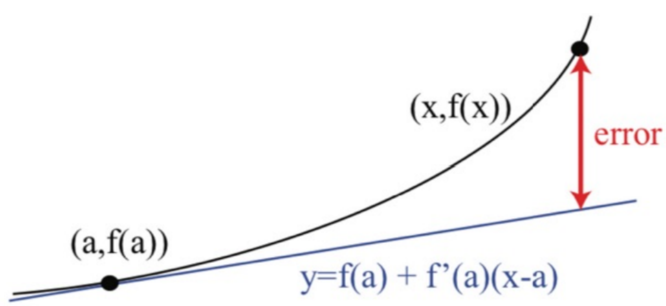


Figure: MVT vs Linear approximation

The MVT tells us that the slope of the secant line from $(a, f(a))$ to $(x, f(x))$ is between the minimum and maximum values of $f'(x)$ on the interval. Linear approximation assumes that $f'(a)$ remains relatively constant, at least until $(xf(x))$. Together this assures us that since Linear approximation projects in a reasonable direction, it gives a reasonable approximation.

A.2.10 Differentials

Here we define the notation of a differential. Given a function $y = f(x)$, the *differential* of y is

$$dy = f'(x)dx$$

Because $y = f(x)$ this can also be called the differential of f. Both dy and $f'(x)dx$ are differentials. The common notation $\frac{dy}{dx} = f'(x)$ can be thought of as the quotient of differentials.

A.3 Integration

A.3.1 Antiderivatives/Indefinite Integrals

Here we introduce the notion of the antiderivative; we say that $G(x) = \int g(x)dx$ is the *antiderivative* of g. Other ways of saying this are:

$$G'(x) = g(x) \quad \text{or,} \quad dG = g(x)dx$$

Note that the definition includes a differential dx. The antiderivative

$$G(x) = \int g(x)dx$$

is also known as the *indefinite integral* of g.

Example: $\sin x$
The integral/antiderivative of $g(x) = \sin x$ is the function whose derivative is $\sin x$. Since the derivative of $-\cos x$ is $\sin x$, it is an antiderivative, $G(x)$ of $\sin x$:

$$\begin{aligned} G(x) &= -\cos x, \quad \text{then} \\ G'(x) &= \sin(x) = g(x) \end{aligned}$$

On the other hand any $G(x) = -\cos x + c$ for any constant c would be valid since the derivative of a constant is 0. Therefore we write:

$$\int \sin x \, dx = -\cos x + c$$

This is called the *indefinite integral* of $\sin x$ because c can be any constant - an indefinite value.

A.3.2 Antiderivative of x^a

Here we describe the antiderivative of x^a, which intuitively would be

$$\text{Since} \quad d(\frac{x^{a+1}}{a+1}) = x^a dx$$
$$\frac{x^{a+1}}{a+1} + c = \int x^a dx \quad \text{for } a \neq -1$$

Note that this antiderivative doesn't work in the case where $a = -1$. Therefore we separately consider a case for the antiderivative of x^-1 or $\frac{1}{x}$, which intuitively is:

$$\int \frac{1}{x} dx = \ln x + c$$

However since $\ln x$ is not valid for $x < 0$, a more standard form would be:

$$\int \frac{1}{x} dx = \ln |x| + c$$

This can be verified by taking the value of $\ln |x|$; since $|x| = x$ doesn't change for $x \geq 0$, we only need to check the case where x is negative.

$$\begin{aligned}
\frac{d}{dx} \ln |x| &= \frac{d}{dx} \ln(-x) \quad (|x| = -x \text{ when } x < 0) \\
&= \frac{1}{-x} \frac{d}{dx}(-x) \quad \text{(chain rule)} \\
&= -\frac{1}{-x} \\
&= \frac{1}{x}
\end{aligned}$$

$\ln |x| + c$ is the antiderivative/indefinite integral of $\frac{1}{x}$ or, in this context, x^a when $a = -1$.

A.3.3 Integration by Substitution with examples

In some cases a 'messy' integral can be substituted with another variable to greatly simplify the integral. This is best illustrated through examples.

Example: $\int x^3(x^4+2)^5 dx$
Integration by substitution involves substituing the 'messiest' function in the integral, in this case we substitute $u = x^4 + 2$. Note that the differential of u is $du = u' dx = 4x^3 dx$. This gives us

$$\begin{aligned}\int x^3(x^4+2)^5 dx &= \int \underbrace{(x^4+2)^5}_{u^5}\underbrace{x^3 dx}_{\frac{1}{4}du} \\ &= \int \frac{u^5}{4} du \\ &= \frac{u^6}{4\cdot 6} + c = \frac{u^6}{24} + c\end{aligned}$$

Notice that the integral is greatly simplified by the introduction of u. Now we can bring plug x back in to get

$$\begin{aligned}\int x^3(x^4+2)^5 dx &= \frac{u^6}{24} + c \\ &= \frac{(x^4+2)^6}{24} + c\end{aligned}$$

Example: $\int \frac{x}{\sqrt{1+x^2}} dx$
We substitute:

$$u = 1 + x^2 \quad \text{and} \quad du = 2x\, dx$$

To get

$$\begin{aligned}\int \frac{x}{\sqrt{1+x^2}} dx &= \int \frac{1}{2\sqrt{u}} du \\ &= \int \frac{1}{2} u^{-1/2} du \\ &= \frac{1}{2} \cdot 2u^{1/2} + c \\ &= u^{1/2} + c \\ &= \sqrt{1+x^2} + c\end{aligned}$$

A.3.4 Intuition for Definite Integrals

Definition
A definite integral can be thought of as the area under a curve, or more specifically, the cumulative sum of the area under a range of discreticised, infinitesimal points on a curve.

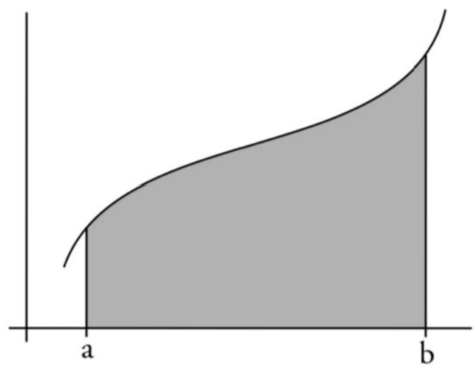


Figure: Area under a curve

The notation we use to describe this is the *definite integral*

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

Unlike an *indefinite* integral, a definite integral has specified start and end points. The idea of discretised rectangles is illustrated here:

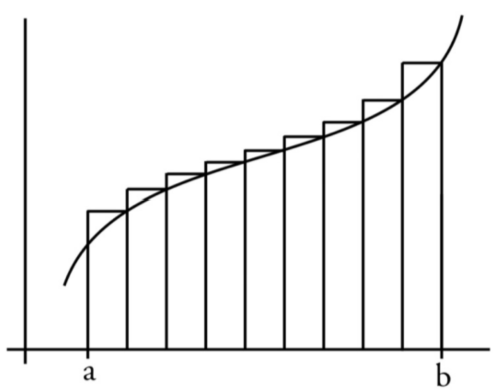


Figure: Area under curve divided into rectangles

The idea here is to take the cumulative area of the rectangles as their width becomes infinitesimally small.
(next page)

Intuition
Consider the curve $f(x) = x^2$, where we add up the area under the curve from point $a = 0$ to point b:

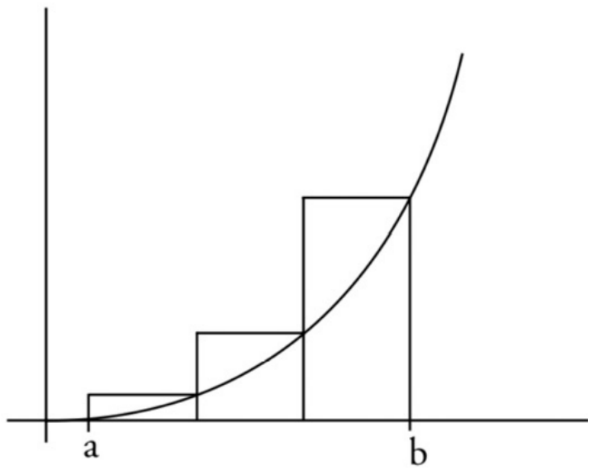


Figure: $\int_a^b f(x)dx$ approximated by 3 rectangles

The interval from $a = 0$ to b is subdivided into n sections; the above figure illustrates $n = 3$. Each subdivision forms the base of a rectangle, with the top of each rectangle touching the graph; here we choose to have the intersection of each rectangle and the graph be at the upper right hand of the rectangle. This causes the cumulative area of the rectangles to be larger than the actual area under the curve.

Now we add up the areas of the rectangles to get an approximation of the area under the curve:

$$\underbrace{\left(\frac{b}{n}\right)}_{\text{base}}\underbrace{\left(\frac{b}{n}\right)^2}_{\text{height}} + \left(\frac{b}{n}\right)\left(\frac{2b}{n}\right)^2 + \left(\frac{b}{n}\right)\left(\frac{3b}{n}\right)^2 + \ldots + \left(\frac{b}{n}\right)\left(\frac{nb}{n}\right)^2$$

First we simplify by factoring out $\left(\frac{b}{n}\right)^3$:

$$\left(\frac{b}{n}\right)\left(\frac{b}{n}\right)^2 + \left(\frac{b}{n}\right)\left(\frac{2b}{n}\right)^2 + \left(\frac{b}{n}\right)\left(\frac{3b}{n}\right)^2 + \ldots + \left(\frac{b}{n}\right)\left(\frac{nb}{n}\right)^2$$
$$= \left(\frac{b}{n}\right)^3 (1 + 2^2 + \ldots + (n-1)^2 + n^2)$$

Next we use geometry to simplify this
(next page)

Consider a pyramid made up of a n by n by 1 base, followed by a $(n-1)$ by $(n-1)$ by 1 layer, and so on until it has a height n.

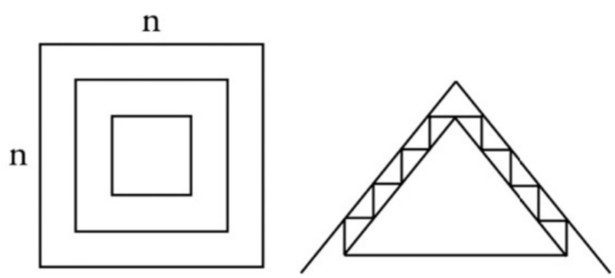


Figure: Top and side views of proposed pyramid

The volume of the pyramid is $(1^2 + 2^2 + \ldots + (n-1)^2 + n^2)$. Now consider the largest pyramid that can fit inside our proposed structure; it would have a volume of $\frac{1}{3} \cdot$ base $\cdot$ height$=\frac{1}{3}n^3$. This gives us the inequality

$$\frac{1}{3}n^3 < 1^2 + 2^2 + \ldots + (n-1)^2 + n^2$$

Finally, consider the the smallest ordinary pyramid that could contain our proposed structure; it would have a base and height of $n+1$, giving it a volume of $\frac{1}{3}(n+1)^3$, completing our inequality:

$$\frac{1}{3}n^3 < 1^2 + 2^2 + \ldots + (n-1)^2 + n^2 < \frac{1}{3}(n+1)^3$$

Dividing throughout by n^3 gives us

$$\frac{1}{3} < \frac{1^2 + 2^2 + \ldots + (n-1)^2 + n^2}{n^3} < \frac{1}{3} \cdot \left(\frac{n+1}{n}\right)^3 = \frac{1}{3}\left(1 + \frac{1}{n}\right)^3$$

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

$$\lim_{n\to\infty} \frac{1^2 + 2^2 + \ldots + (n-1)^2 + n^2}{n^3} = \frac{1}{3}$$

Thus, going back to our original problem:

$$\left(\frac{b}{n}\right)^3 (1 + 2^2 + \ldots + (n-1)^2 + n^2) = b^3 \left(\frac{1 + 2^2 + \ldots + (n-1)^2 + n^2}{n^3}\right) = \frac{1}{3}b^3$$

This leads us to the result:

$$\int_0^b x^2 dx = \frac{1}{3}b^3$$

A.3.5 Riemann Sums

Here we define a Riemann Sum. An intuitive way to find a definite integral $\int_a^b f(x)\,dx$ is by approximating the area under a curve using sums of rectangles.

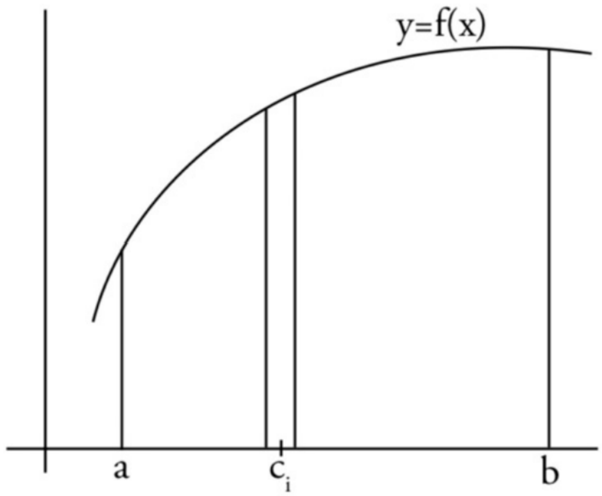


Figure: Area under a curve

With this we can come up with a general procedure:

- Divide $[a, b]$ into n equal pieces of length $\Delta x = \frac{b-a}{n}$.
- Pick *any* c_i in the i^{th} interval and use $f(c_i)$ to compute the height for each rectangle.
- Sum the areas of the rectangles:

$$\underbrace{f(c_1)}_{\text{height}}\underbrace{(\Delta x)}_{\text{base}} + f(c_2)\Delta x + \ldots + f(c_{n-1})\Delta x + f(c_n)\Delta x$$

$$= \sum_{i=1}^{n} f(c_i)\Delta x$$

The sum $\sum_{i=1}^{n} f(c_i)\Delta x$ is called a **Riemann Sum**.

As n approaches infinity, the sum approaches the value of a definite integral:

$$\sum_{i=1}^{n} f(c_i)\Delta x = \int_a^b f(x)\,dx$$

Which is the area under the curve $f(x)$ above the interval $[a, b]$.

A.3.6 Introduction to Fundamental theorem of Calculus (FTC1)

Here we introduce the fundamental theorem of calculus. There are two versions of it so we abbreviate them as FTC1 and FTC2. Here we cover FTC1.

Theorem: If $f(x)$ is continuous and $F'(x) = f(x)$, then:

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

Recall the notation of the antiderivative:

$$F(x) = \int f(x)\,dx$$

The FTC1 can also be written as

$$\int_a^b f(x)\,dx = F(x)|_a^b$$

Example: Consider $f(x) = x^2$, using FTC we get:

$$\int_a^b x^2 dx = \int_a^b f(x)\,dx = F(b) - F(a) = \frac{b^3}{3} - \frac{a^3}{3}$$

By using the FTC we avoid the elaborate computation required by Riemann sums.

A.3.7 Properties of Integrals

Here we list a number of properties of integrals. Most are intuitive:

1. The integral of a sum is the sum of integrals

$$\int_a^b (f(x) + g(x))\, dx = \int_a^b f(x)\, dx + \int_a^b g(x)\, dx$$

2. Constant multiples can be factored out:

$$\int_a^b cf(x)\, dx = c \int_a^b f(x)\, dx$$

(scaling the area under a graph by a multiple is the same as scaling the graph itself by the multiple then finding the area under it)

3. We can combine definite integrals; if $a < b < c$ then

$$\int_a^b f(x)\, dx + \int_b^c f(x)\, dx = \int_a^c f(x)\, dx$$

Graphically:

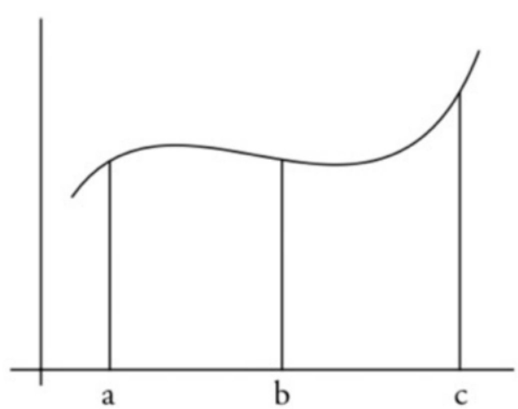


Figure: Combining two areas under a curve

4. A definite integral over a range of length 0 is 0:

$$\int_a^a f(x)\, dx = 0$$

(next page)

5. Taking the upper limit to be lower than the lower limit:

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

(Since $F(b) - F(a) = -(F(a) - F(b))$)

6. *Estimation*: If $f(x) \leq g(x)$ and $a < b$, then

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

7. *Substitution/Change of variables*: In terms of indefinite integrals, if $u = u(x)$ then $du = u'(x)\,dx$ and $\int g(u)\,du = \int g(u(x))u'(x)\,dx$ (substitution). We can apply this concept to definite integrals by adapting the limits accordingly:

$$\int_{u_1}^{u_2} g(u)\,du = \int_{x_1}^{x_2} g(u(x))u'(x)\,dx$$

where $u_1 = u(x_1)$ and $u_2 = u(x_2)$. Note that this is true if u is always increasing or always decreasing on $x_1 < x < x_2$ (the variables u and x must be one-to-one or a value of u can correspond to multiple values of x or vice-versa.)

Example of Substitution/Change of variables:
Consider $\int_1^2 (x^3+2)^5 x^2 dx$, we attempt to find the result by substitution. Here we use $u = x^3 + 2$, taking the limits into account; we have $du = 3x^2 dx$, $u(1) = 3$ and $u(2) = 10$. This gives us:

$$\begin{aligned}
\int_{x=1}^{x=2} (x^3+2)^5 x^2 dx &= \int_{u=3}^{u=10} u^5 \frac{1}{3} du \\
&= \left.\frac{u^6}{6 \cdot 3}\right|_{u=3}^{u=10} \\
&= \frac{1}{18}(10^6 - 3^6)
\end{aligned}$$

A.3.8 The Fundamental Theorem and the Mean Value Theorem

Here we compare the FTC1 with the Mean Value Theorem (MVT). Substituting $\Delta F = F(b) - F(a)$ and $\Delta x = b - a$, the FTC1 tells us that:

$$\Delta F = \int_a^b f(x)\,dx$$

Dividing both sides by Δx:

$$\frac{\Delta F}{\Delta x} = \underbrace{\frac{1}{b-a}\int_a^b f(x)\,dx}_{\text{Average}(f)}$$

Consider the Riemann sum:

$$\int_0^n f(x)\,dx \approx f(1) + f(2) + \ldots + f(n)$$

Which is a cumulative sum of the values of $f(x)$ over a specified range; this means that

$$\begin{aligned}\frac{\int_0^n f(x)\,dx}{n} &\approx \frac{f(1) + f(2) + \ldots + f(n)}{n} \\ &= \text{Average}(f) \quad \text{(Over the specified range)}\end{aligned}$$

Going back to the original problem, this means that:

$$\begin{aligned}\frac{\Delta F}{\Delta x} &= \frac{1}{b-a}\int_a^b f(x)\,dx \\ &= \text{Average}(F') \quad \text{(Over a range from } a \text{ to } b\text{)}\end{aligned}$$

Rearranging:

$$\Delta F = \text{Average}(F')\Delta x$$

Compare this with the MVT, which gives us

$$\begin{aligned}\frac{\Delta F}{\Delta x} = \frac{F(b) - F(a)}{b-a} &= F'(c) \quad \text{For some } c \text{ where } a \le c \le b \\ \Delta F &= F'(c)\Delta x\end{aligned}$$

(next page)

The FTC1 and MVT give us:

$$\begin{aligned} \text{MVT:} \quad & \Delta F = F'(c)\Delta x \\ \text{FTC1:} \quad & \Delta F = \text{Average}(F')\Delta x \end{aligned}$$

The value of Average(F') in the first fundamental theorem is very specific, while $F'(c)$ from the mean value theorem is not, since all we know is $a \le c \le b$.

For the MVT, even though we don't know what c is, we can say that

$$\left(\min_{a<x<b} F'(x)\right) \le F'(c) \le \left(\max_{a<x<b} F'(x)\right)$$
$$\left(\min_{a<x<b} F'(x)\right)\Delta x \le F'(c)\Delta x = \Delta F \le \left(\max_{a<x<b} F'(x)\right)\Delta x$$

Notice that the FTC1 allows us to make the same deduction, but with a much more specific function Average(F')

$$\left(\min_{a<x<b} F'(x)\right) \le \text{Average}(F') \le \left(\max_{a<x<b} F'(x)\right)$$
$$\left(\min_{a<x<b} F'(x)\right)\Delta x \le \text{Average}(F')\Delta x = \Delta F \le \left(\max_{a<x<b} F'(x)\right)\Delta x$$

The fundamental theorem of calculus is clearly much stronger than the MVT, allowing us to abandon the mean value theorem, since either theorem gives us the same conclusion

$$\left(\min_{a<x<b} F'(x)\right)\Delta x \le \Delta F \le \left(\max_{a<x<b} F'(x)\right)\Delta x$$

A.3.9 Second Fundamental Theorem of Calculus

Here we introduce the Second Fundamental Theorem of Calculus. These two theorems are key to finding the intuition toward the relationship between integrals and derivatives.

Theorem: If f is continuous and $G(x) = \int_a^x f(t)\,dt$, then $G'(x) = f(x)$.

Intuition
The variable t is a dummy variable. What matters here is the upper limit—x. The upper limit of the integral is *changing with respect to* x. This is the link between the area under a curve (integrals) and the change with respect to x (derivatives). See below (A.3.10).

A.3.10 Intuition and proofs—Fundamental Theorem of Calculus

Second Fundamental Theorem
Recall the FTC2: If f is continuous and $F(x) = \int_a^x f(t)\,dt$, then $F'(x) = f(x)$. Here we use the interpretation that $F(x)$ equals the area under the curve from a to x, taking its derivative to show that its equal to f. This links the idea of the integral and the derivative.

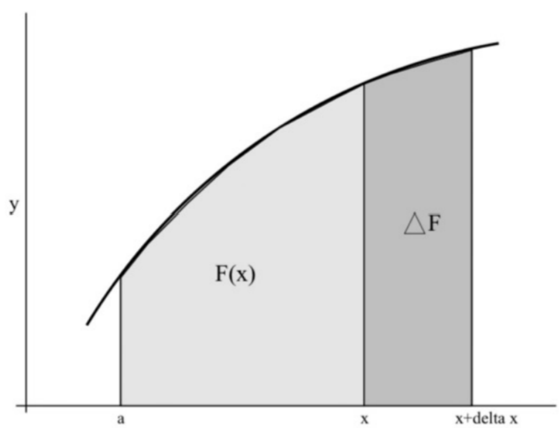


Figure: Graph of $f(x)$ with shaded area $F(x)$

As Δx becomes smaller, we can apporximate ΔF to be a rectangle of width Δx and height $f(x)$. So

$$\Delta F \approx \Delta x f(x)$$

So as $\Delta x \to 0$:

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

(next page)

First Fundamental Theorem

Theorem: (First Fundamental Theorem of Calculus) If f is continuous and $F' = f$, then $\int_a^b f(x)\,dx = F(b) - F(a)$.

Here we define an antiderivative G of f before using it to calculate $F(b) - F(a)$. We start with $F' = f$, where f is continuous (not strictly necessary, but allows us to use FTC2). Now consider $G(x) = \int_a^x f(t)\,dt$; by FTC2:

$$G'(x) = f(x)$$

So $F'(x) = G'(x) = f(x)$. This means that F and G differ only by a constant:

$$F(x) = G(x) + c$$

Now we can show:

$$\begin{aligned} F(b) - F(a) &= (G(b) + c) - (G(a) + c) \\ &= G(b) - G(a) \\ &= \int_a^b f(t)\,dt - \int_a^a f(t)\,dt \\ &= \int_a^b f(t)\,dt - 0 \\ F(b) - F(a) &= \int_a^b f(x)\,dx \end{aligned}$$

A.3.11 Integrals and Averages

Here we show how integrals can act as continous analogues of averages. Consider finding the average value of a function $y - f(x)$ on an interval:

$$\text{Discrete Average} \approx \frac{y_1 + y_2 + \ldots + y_n}{n}$$
$$\text{Continuous Average} = \frac{1}{b-a}\int_a^b f(x)\,dx = \text{Ave}(f)$$

The continous average is essentially the discrete average as the number of values n approaches infinity.

Intuition
Consider $f(x)$ over an interval $[a, b]$:

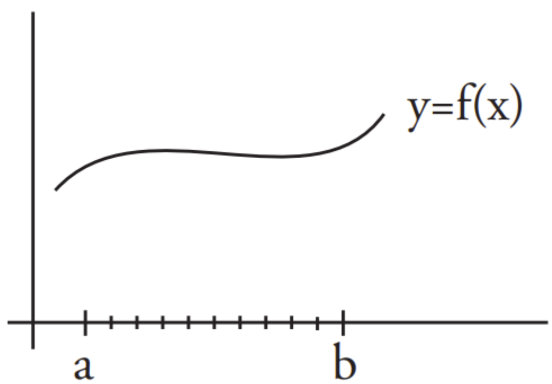


Figure: $f(x)$ over $[a, b]$

Consider $n + 1$ equally spaced points where $a = x_0 < x_1 < \ldots < x_n = b$ with the distance between each point being $\Delta x = \frac{b-a}{n}$. For $y_i = f(x_i)$, the Riemann sum approximating the area under f over the specified range would be

$$(y_0 + y_1 + \ldots + y_n)\Delta x$$

As $n \to \infty$, the Riemann sum approaches the area under the curve,

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

(next page)

Now we can show, as $n \to \infty$

$$\begin{aligned}
\int_a^b f(x)\,dx &\approx (y_0 + y_1 + \ldots + y_n)\Delta x \\
\frac{1}{b-a}\int_a^b f(x)\,dx &\approx \frac{1}{b-a}(y_0 + y_1 + \ldots + y_n)\Delta x \\
&= \frac{1}{b-a}(y_0 + y_1 + \ldots + y_n)\frac{b-a}{n} \\
\frac{1}{b-a}\int_a^b f(x)\,dx &\approx \frac{y_0 + y_1 + \ldots + y_n}{n}
\end{aligned}$$

Weighted averages
Similarly, continuous *weighted averages* are given by

$$\frac{\int_a^b f(x)w(x)\,dx}{\int_a^b w(x)\,dx}$$

Where a discrete analogue might look like

$$\frac{10w_1 + 20w_2 + 30w_3}{w_1 + w_2 + w_3}$$

A.3.12 Introduction—Numerical Integration

Many functions may not have easily describable antiderivatives, so many integrals must be approximated by a computer/calculator. The simplest of these techniques is Riemann Sums:

Riemann Sums in numerical integration
Riemann Sums approximate the area between the area between the x-axis and the curve over the interval $[a, b]$ by a sum of areas of rectangles:

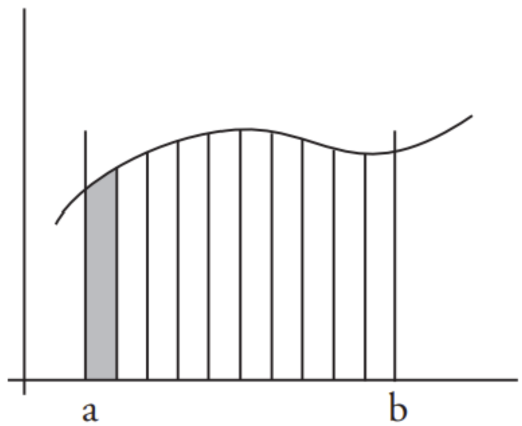


Figure: Riemann sum of $f(x)$ over $[a, b]$

Each rectangle has a width $\Delta x = x_i - x_{i-1}$. There are n rectangles whose top edges have x-coordinates $a = x_0 < x_1 < \ldots < x_n = b$; the heights of the rectangles are therefore $y_0 = f(x_0), y_1 = f(x_1), \ldots, y_n = f(x_n)$. Note that one can use either left or right Riemann sums to approximate the area. The formula for the left Riemann sum, where the top left corner of each rectangle is set at the height of the graph, is

$$(y_0 + y_1 + \ldots + y_{n-1})\Delta x$$

Similarly, if we left the top right corner of each rectangle be set to the height of the graph we get the right Riemann sum:

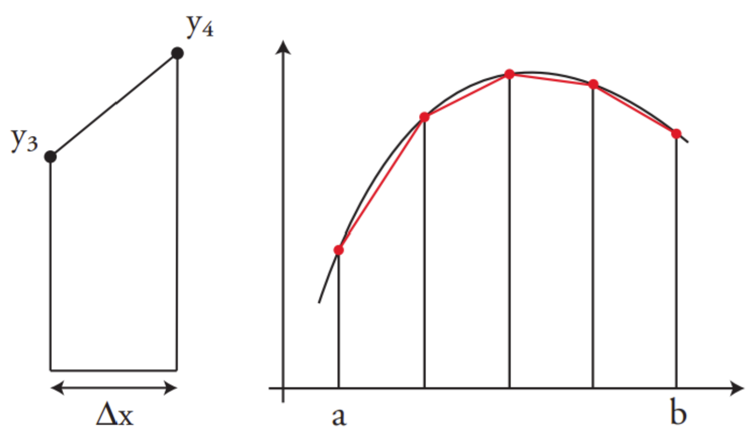
$$(y_1 + y_2 + \ldots + y_n)\Delta x$$

In the end both act as (inefficient) approximations to

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

A.3.13 Numerical Integration—Trapezoidal Rule

Similar to Numerical integration by Riemann sums, the Trapezoidal rule divides the area under the graph into trapezoids (using segments of secant lines) rather than rectangles (Riemann sums).



Left: Single trapezoid

Right: Trapezoids approximate integral more accurately

The trapezoids are able to approximate the area under the curve more accurately. Each trapezoid has a width $\Delta x = x_i - x_{i-1}$ where there are n trapezoids whose bottom edges have x-coordinates $a = x_0 < x_1 < \ldots < x_n = b$. The heights of each trapezoid are y_{i-1} and y_i where $y_i = f(x_i)$.

Trapezoid i has area

$$\Delta x \left(\frac{y_{i-1} + y_i}{2} \right)$$

Adding up the areas of all the trapezoids we get

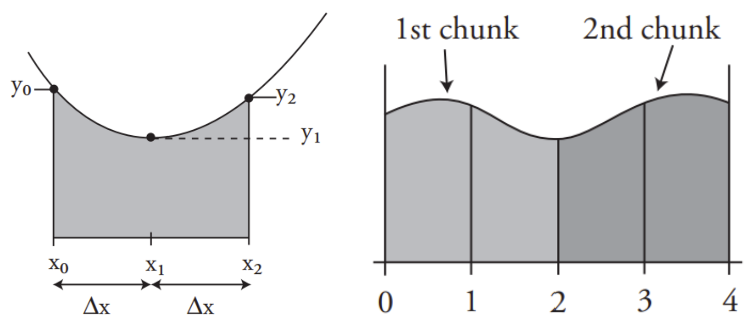
$$\begin{aligned} \text{Area} &= \Delta x \left(\frac{y_0 + y_1}{2} + \frac{y_1 + y_2}{2} + \ldots + \frac{y_{n-1} + y_n}{2} \right) \\ &= \Delta x \left(\frac{y_0}{2} + y_1 + y_2 + \ldots + y_{n-1} + \frac{y_n}{2} \right) \end{aligned}$$

Notice that the trapezoidal rule is the average of the left and right Riemann sums. It is much better than a Riemann sum, but is still not very efficient.

A.3.14 Numerical Integration—Simpson's rule

Intuition:
Here we introduce Simpson's rule, which is similar to Riemann sum and Trapezoid rule approximation, but instead approximates $f(x)$ using parabolas instead of horizontal or secant lines.

Same with the former two approximations, we divide the range we want to approximate into n intervals, with the base of each interval being $\Delta x = x_i - x_{i-1}$. We use the corresponding $y_i = f(x_i)$ to fit a parabola to the curve at each interval.



Left: Parabolic approximation over two intervals/one chunk
Right: Simpson's rule for n intervals (n must be even)

Notice we fit a parabola over an interval of $2\Delta x$, and therefore require an *even* number of intervals over the range we are approximating. Each 'chunk' of two intervals has the area:

$$\frac{\Delta x}{3}(y_0 + 4y_1 + y_2)$$

The area of all n chunks summed up gives us the total estimate of the area:

$$\int_a^b f(x)\,dx \approx \frac{\Delta x}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \ldots + 2y_{n-2} + 4y_{n-1} + y_n)$$

A.3.15 Simpson's rule—Derivation

Here we derive the formula for Simpson's rule, a method for Numerical integration:

$$\int_a^b f(x)\,dx \approx \frac{\Delta x}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \ldots + 2y_{n-2} + 4y_{n-1} + y_n)$$

The area under our function is divided into an even number of intervals as shown, approximation occurs in 'chunks', which is equivalent to 2 intervals:

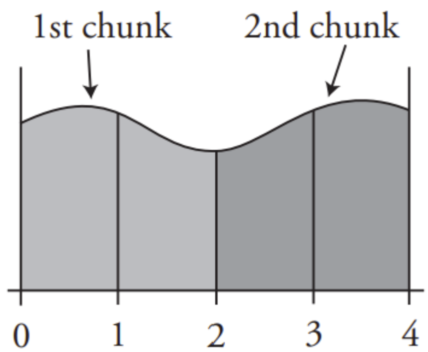


Figure: Each chunk is made up of two intervals

First we derive the equation for the area of each 'chunk' approximating the area of one of n segments:

$$\frac{\Delta x}{3}(y_0 + 4y_1 + y_2)$$

Each 'chunk' is equal to the width of two intervals/$2\Delta x$:

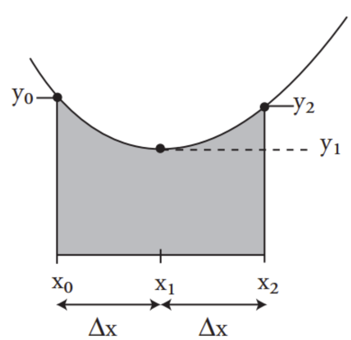


Figure: We use three points to approximate a parabola over two Δx

(Notice that we fit a parabola over an interval of $2\Delta x$, and therefore require an *even* number of intervals over the range we are approximating.)
(next page)

Our approximation $P(x)$ for the upper edge of our area has the formula:

$$P(x) = ax^2 + bx + c$$

Thus we require three points. Taking x_1 to be 0 (this can be done since we only care about estimating the area below a parabola made of our three y_i), our area can be expressed as:

$$\text{Area} = \int_{-\Delta x}^{\Delta x} P(x)\,dx = \int_{-\Delta x}^{\Delta x} ax^2 + bx + c\,dx$$

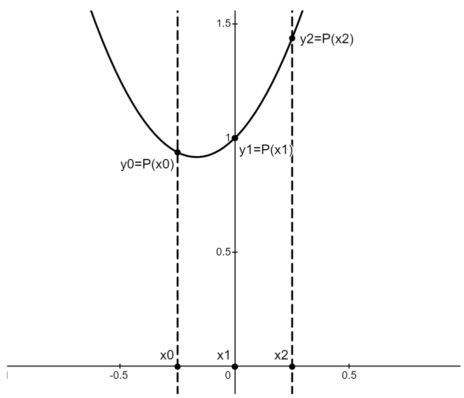


Figure: Our points shifted such that $x_1 = 0$

Simplifying our integral:

$$\begin{aligned}\text{Area} = \int_{-\Delta x}^{\Delta x} ax^2 + bx + c\,dx &= \frac{a}{3}x^3 - \frac{b}{2}x^2 + cx\Big|_{-\Delta x}^{\Delta x} \\ &= \frac{2}{3}a(\Delta x)^3 + 2c\Delta x\end{aligned}$$

Since we have $x_1 = 0$, $P(x_1) = P(0) = y_1 = c$. This means

$$\begin{aligned}\text{Area} = \frac{2}{3}a(\Delta x)^3 + 2c\Delta x &= \frac{2}{3}a(\Delta x)^3 + 2y_1\Delta x \\ &= \frac{\Delta x}{3}(2a(\Delta x)^2 + 6y_1)\end{aligned}$$

(next page)

Since

$$\begin{aligned} P(\Delta x) &= a(\Delta x)^2 + b\Delta x + c = y_2 \\ P(-\Delta x) &= a(\Delta x)^2 - b\Delta x + c = y_0 \end{aligned}$$

and $y_1 = c$, we can also show

$$\begin{aligned} 2a(\Delta x)^2 + 2c &= y_0 + y_2 \\ 2a(\Delta x)^2 &= y_0 + y_2 - 2y_1 \end{aligned}$$

Therefore, returning to finding the area under our single chunk:

$$\begin{aligned} \text{Area} &= \frac{\Delta x}{3}(2a(\Delta x)^2 + 6y_1) \\ &= \frac{\Delta x}{3}(y_0 + y_2 - 2y_1 + 6y_1) \\ &= \frac{\Delta x}{3}(y_0 + 4y_1 + y_2) \end{aligned}$$

We add up the chunks to get the approximated value of our integral:

$$\text{Total Area} = \frac{\Delta x}{3}[(y_0 + 4y_1 + y_2) + (y_2 + 4y_3 + y_4) + \ldots + (\ldots y_n)]$$

With that we get the final form of Simpson's rule

$$\int_a^b f(x)\,dx \approx \frac{\Delta x}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + \ldots + 2y_{n-2} + 4y_{n-1} + y_n)$$

A.3.16 Area under Bell Curve

Here we prove the value

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$$

Volume about vertical axis:
Consider the volume of revolution created by rotating the curve e^{-r^2} around the *vertical* axis:

$$V = \int_0^\infty 2\pi r e^{-r^2} dr$$

(Think of it as a Riemann sum of circular 3-dimensional strips around the vertical axis, with each strip having circumference $2\pi r$, height e^{-r^2} and thickness dr). This evaluates to:

$$\begin{aligned} V &= \int_0^\infty 2\pi r e^{-r^2} dr \\ &= -\pi e^{-r^2}\Big|_0^\infty \\ &= \pi \end{aligned}$$

Area under curve:
We aim to find the volume under the entire curve:

$$Q = \int_{-\infty}^\infty e^{-t^2} dt$$

Illustrated:

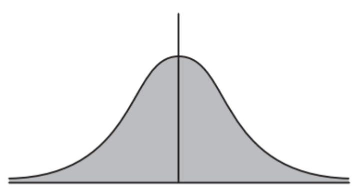


Figure: Q=Area under e^{-t^2}

Now we show that

$$\begin{aligned} Q^2 &= V = \pi \\ Q &= \sqrt{\pi} \end{aligned}$$

(next page)

We prove that $V = Q^2$ using slices; imagine 3-dimsntional slices of our revolution of e^{-r^2} about the vertical axis:

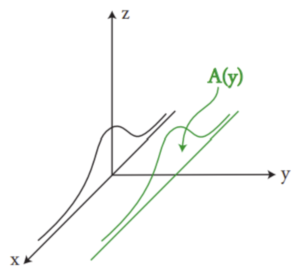


Figure: Three dimensional slices of the volume of rotation of e^{-r^2}

Where each slice has an area $A(y)$. Here we slice along the y-axis. We want an expression for $A(y)$, so we fix $y = b$ and vary x, using the pythagorean theorem to find r:

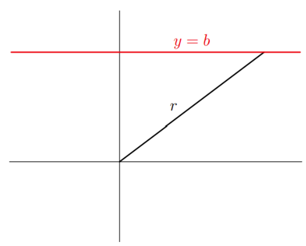


Figure: Top down view of a single slice; $r^2 = b^2 + x^2$

The height of our slice follows the formula

$$\text{Height} = e^{-r^2} = e^{-(b^2+x^2)}$$

The total volume of our slices is equal to V, and can be expressed as

$$V = \int_{-\infty}^{\infty} A(y)\,dy$$

With our y-coordinate fixed at b,

$$\begin{aligned} A(b) &= \int_{-\infty}^{\infty} e^{-(b^2+x^2)} dx \\ &= e^{-b^2} \int_{-\infty}^{\infty} e^{-x^2} dx \\ &= e^{-b^2} Q \end{aligned}$$

(next page)

Since

$$A(b) = e^{-b^2} Q$$

And our volume is given by

$$V = \pi = \int_{-\infty}^{\infty} A(y)\, dy$$

We can show

$$\begin{aligned} V &= \int_{-\infty}^{\infty} A(y)\, dy \\ &= \int_{-\infty}^{\infty} e^{-y^2} Q\, dy \\ &= Q \int_{-\infty}^{\infty} e^{-y^2}\, dy \\ &= Q^2 \quad (\text{since } Q = \int_{-\infty}^{\infty} e^{-t^2} dt) \end{aligned}$$

With that we get

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-t^2} dt &= Q \\ &= \sqrt{V} \\ &= \sqrt{\pi} \\ \int_{0}^{\infty} e^{-t^2} dt &= \frac{\sqrt{\pi}}{2} \end{aligned}$$

A.4 Techniques of Integration

A.4.1 Review of Trigonometric Identities ($\sin$ and $\cos$)

Here we review angle formulas, in particular

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

Recall:

$$\sin^2\theta + \cos^2\theta = 1$$
$$\cos(2\theta) = \cos^2\theta - \sin^2\theta$$
$$\sin(2\theta) = 2\sin\theta\cos\theta$$

We can show

$$\begin{aligned}\cos(2\theta) &= \cos^2\theta - \sin^2\theta \\ &= \cos^2\theta - (1 - cos^2\theta) \\ &= 2\cos^2\theta - 1 \\ \cos^2\theta &= \frac{1+\cos(2\theta)}{2}\end{aligned}$$

We can further show

$$\begin{aligned}\sin^2\theta &= 1 - \cos^2\theta \\ &= 1 - \left(\frac{1+\cos(2\theta)}{2}\right) \\ &= \frac{1-\cos(2\theta)}{2}\end{aligned}$$

A.4.2 Integral of $\sin^n x \cos^m x$—Part 1 (Odd Exponents)

Here we derive more complicated formulas involving integrals of the form:

$$\int \sin^n x \cos^m x \, dx$$

Where m and n are non-negative integers. The integral can be divided into two cases. First we consider the easier case—where at least one exponent is odd.

Example: $m = 1$, $\int \sin^n x \cos x$
The idea here is to integrate via substitution, where $u = \sin x$ so $du = \cos x \, dx$:

$$\begin{aligned}\int \sin^n x \cos x \, dx &= \int u^n \, du \\ &= \frac{u^n + 1}{n+1} + c \\ &= \frac{\sin^{n+1} x}{n+1} + c\end{aligned}$$

Example: $\int \sin^3 x \cos^2 x \, dx$
Similar to the first example, one of the exponents is odd. We turn this integral into one in which the odd exponent is 1 using trigonometric identities. In this case the odd exponent is on $\sin x$, so we use

$$\sin^2 x = 1 - \cos^2 x$$

so we get

$$\begin{aligned}\int \sin^3 x \cos^2 x \, dx &= \int (1 - \cos^2 x) \cdot \sin x \cos^2 x \, dx \\ &= \int (\cos^2 x - \cos^4 x) \cdot \sin x \, dx\end{aligned}$$

Similarly, we substitute the $u = \cos x$, $du = -\sin x \, dx$ (notice this removes the single exponent function):

$$\begin{aligned}\int (\cos^2 x - \cos^4 x) \cdot \sin x \, dx &= \int (u^4 - u^2) \, du \\ &= \frac{u^5}{5} - \frac{u^3}{3} + c \\ &= \frac{\cos^5 x}{5} - \frac{\cos^3 x}{3} + c\end{aligned}$$

A.4.3 Integral of $\sin^n x \cos^m x$—Part 2 (Even Exponents)

Here we consider the harder case— $\int \sin^n x \cos^m x \, dx$ where both exponents are even. This case is harder by virtue of being more tedious; it is solved by repeatedly applying angle formulas rather than integration by substitution.

Example: $\int \cos^2 x \, dx$
This example illustrates that the antiderivative for $\cos^2 x$ is not straight forward:

$$\begin{aligned}\int \cos^2 x \, dx &= \int \frac{1+\cos(2x)}{2} dx \\ &= \frac{x}{2} - \frac{\sin(2x)}{4} + c\end{aligned}$$

Example: $\int \sin^2 x \cos^2 x \, dx$
We use angle formulas to turn our *product* of functions into *sums*:

$$\begin{aligned}\sin^2 x \cos^2 x \, dx &= \left(\frac{1-\cos(2x)}{2}\right) \cdot \left(\frac{1+\cos(2x)}{2}\right) \\ &= \frac{1-\cos^2(2x)}{4} \\ &= \frac{1}{4} - \frac{(1+\cos(4x))/2}{4} \\ &= \frac{1}{8} - \frac{\cos(4x)}{8}\end{aligned}$$

Our integral therefore looks like:

$$\begin{aligned}\int \sin^2 x \cos^2 x \, dx &= \int \frac{1}{8} - \frac{\cos(4x)}{8} dx \\ &= \frac{x}{8} - \frac{\sin(4x)}{32} + c\end{aligned}$$

(next page)

Note that our example $\int \sin^2 x \cos^2 x \, dx$ can also be separated using the other half angle formula

$$\sin(2\theta) = 2 \sin\theta \cos\theta$$

we end up with the same expression

$$\begin{aligned}
\sin^2 x \cos^2 x &= (\sin x \cos x)^2 \\
&= \left(\frac{\sin(2x)}{2} \right)^2 \\
&= \frac{\sin^2(2x)}{4} \\
&= \frac{1}{4} \left(\frac{1 - \cos(4x)}{2} \right) \\
&= \frac{1}{8} - \frac{\cos(4x)}{8}
\end{aligned}$$

A.4.4 Review of Trigonometric identities (tan, sec)

Here we briefly show

$$\sec^2 x = 1 + \tan^2 x$$

this is apparent once we express sec in terms of sin and cos:

$$\begin{aligned}
\sec^2 x &= \frac{1}{\cos^2 x} \\
&= \frac{\cos^x + \sin^2 x}{\cos^2 x} \\
&= 1 + \tan^2 x
\end{aligned}$$

A.4.5 Integration of tan, sec, csc, cot

Integrating $\tan x$:
Intuitively, we know

$$\int \sec^2 x \, dx = \tan x + c$$

This is relatively simple since it naturally falls out of the derivative of $\tan x$. Here we derive the integral for $\tan x$:

$$\int \tan x \, dx = -\ln|\cos x| + c$$

We express $\tan x$ in terms of $\sin x$ and $\cos x$ before substituting $u = \cos x$, $du = -\sin x \, dx$:

$$\begin{aligned}
\int \tan x \, dx &= \int \frac{\sin x}{\cos x} dx \\
&= \int -\frac{1}{u} du \\
&= -\ln|u| + c \\
&= -\ln|\cos x| + c
\end{aligned}$$

(next page)

Integrating $\cot x$**:**
We show

$$\int \cot x \, dx = -\ln|\sin x| + c$$

Substituting $u = \sin x$, $du = -\cos x \, dx$

$$\begin{aligned}
\int \cot x \, dx &= \int \frac{\cos x}{\sin x} dx \\
&= \int -\frac{1}{u} du \\
&= -\ln|u| + c \\
&= -\ln|\sin x| + c
\end{aligned}$$

Integrating $\sec x$**:**
We show

$$\int \sec x \, dx = \ln|\sec x + \tan x| + c$$

Consider that

$$\begin{aligned}
\frac{d}{dx}(\sec x + \tan x) &= \frac{\sin x}{\cos^2 x} + \sec^2 x \\
&= \sec x \tan x + \sec^2 x \\
&= \sec x \cdot (\sec x + \tan x)
\end{aligned}$$

Substituting $u = \sec x + \tan x$, notice

$$\begin{aligned}
\frac{d}{dx}(\sec x + \tan x) &= \sec x \cdot (\sec x + \tan x) \\
\sec x &= \frac{\frac{d}{dx}(\sec x + \tan x)}{\sec x + \tan x} \\
&= \frac{u'}{u} = (\ln|u|)'
\end{aligned}$$

Therefore we conclude

$$\begin{aligned}
\sec x = (\ln|u|)' &= (\ln|\sec x + \tan x|)' \\
\int \sec x \, dx &= \ln|\sec x + \tan x| + c
\end{aligned}$$

(next page)

Integrating $\csc x$**:**
Here we show

$$\int \csc x \, dx = \ln|\csc x - \cot x| + c$$

First note the derivatives of $\csc x$ and $\cot x$:

$$\begin{aligned}\frac{d}{dx}\csc x &= -\frac{\cos x}{\sin^2 x} \\ &= -\csc x \cot x \\ &\text{and} \\ \frac{d}{dx}\cot x &= \frac{-\sin^2 x - \cos^2 x}{\sin^2 x} \\ &= -\csc^2 x\end{aligned}$$

Proof is similar to that of the integral of $\sec x$:

$$\begin{aligned}\frac{d}{dx}(\csc x - \cot x) &= -\csc x \cot x + \csc^2 x \\ &= \csc x \cdot (\csc x - \cot x)\end{aligned}$$

Where

$$\begin{aligned}\csc x &= \frac{\frac{d}{dx}(\csc x - \cot x)}{(\csc x - \cot x)} \\ &= \frac{d}{dx}\ln|\csc x - \cot x|\end{aligned}$$

Thus we conclude

$$\int \csc x \, dx = \ln|\csc x - \cot x| + c$$

A.4.6 Trigonometric substitution (and Polar coordinates)

Here we introduce trigonometric substitution, which helps solve ugly integrals. Consider:

$$\int \frac{1}{x^2\sqrt{1+x^2}}dx$$

notice this fairly ugly integral can be made simpler by substituting $x = \tan\theta$, $dx = \sec^2\theta\, d\theta$, and $\sec^2 = 1 + \tan^2\theta$:

$$\begin{aligned}\int \frac{1}{x^2\sqrt{1+x^2}}dx &= \int \frac{\sec^2\theta}{\tan^2\theta\sqrt{1+\tan^2\theta}}d\theta \\ &= \int \frac{\sec^2\theta}{\tan^2\theta\sqrt{\sec^2\theta}}d\theta \\ &= \int \frac{\sec\theta}{\tan^2\theta}d\theta \\ &= \int \frac{\frac{1}{\cos\theta}}{\frac{\sin^2\theta}{\cos^2\theta}}d\theta \\ &= \int \frac{\cos\theta}{\sin^2\theta}d\theta\end{aligned}$$

which can be evaluated ($u = \sin\theta$, $du = \cos\theta\, d\theta$):

$$\begin{aligned}\int \frac{1}{x^2\sqrt{1+x^2}}dx &= \int \frac{\cos\theta}{\sin^2\theta}d\theta \\ &= \int \frac{1}{u^2}du \\ &= -\frac{1}{u} + c \\ &= -\csc\theta + c\end{aligned}$$

Now for expression in terms of x, consider our substitution $x = \tan\theta$ visualised:

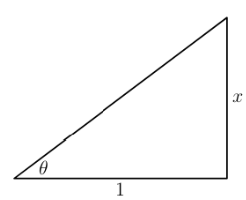


Figure: Undoing trig substitution

This gives us $1/\sin\theta = \csc\theta = \sqrt{1+x^2}/x$ and

$$\int \frac{1}{x^2\sqrt{1+x^2}}dx = -\frac{\sqrt{1+x^2}}{x} + c$$

(next page)

Example: Polar Coordinates
Consider a circle of radius a, cut out a tab of height b:

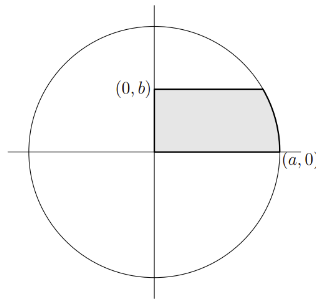


Figure: Notice the upper border is non-differentiable

An intuitive approach might be to integrate with respect to y, where the circle equation in terms of x would be $x = \sqrt{a^2 - y^2}$:

$$\text{Area} = \int_0^b x\,dy = \int_0^b \sqrt{a^2 - y^2}\,dy$$

One way to integrate this is by utilising polar coordinates:

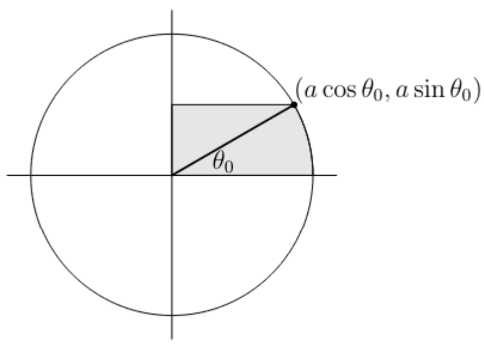


Figure: Expression in terms of trigonometric functions

We can express the coordinates of the upper right corner in terms of *polar coordinates* $(a\cos\theta_0, a\sin\theta_0)$ (this is true in general).
(next page)

Given our integral

$$\text{Area} = \int_0^b \sqrt{a^2 - y^2}\, dy$$

First consider the *indefinite integral*; we substitute $y = a \sin \theta$ and $dy = a \cos \theta \, d\theta$:

$$\begin{aligned}
\int \sqrt{a^2 - y^2}\, dy &= \int x \, dy \\
&= \int (a \cos \theta)(a \cos \theta) d\theta \\
&= a^2 \int \cos^2 x \, d\theta \\
&= a^2 \left(\frac{\theta}{2} + \frac{\sin(2\theta)}{4} \right) + c
\end{aligned}$$

We want to express our equation back in terms of y:

$$\begin{aligned}
\int \sqrt{a^2 - y^2}\, dy &= a^2 \left(\frac{\theta}{2} + \frac{\sin(2\theta)}{4} \right) + c \\
&= a^2 \left(\frac{\theta}{2} + \frac{\sin\theta \cos\theta}{2} \right) + c \\
&= \frac{a^2\theta}{2} + \frac{a \sin\theta \, a \cos\theta}{2} + c
\end{aligned}$$

using $y = a \sin \theta$, we substitute

$$\theta = \arcsin\left(\frac{y}{a}\right)$$

To get

$$\int \sqrt{a^2 - y^2}\, dy = \frac{a^2 \arcsin(y/a)}{2} + \frac{y\sqrt{a^2 - y^2}}{2} + c$$

Our definite integral, and therefore area, is given by

$$\begin{aligned}
\int_0^b \sqrt{a^2 - y^2}\, dy &= \left(\frac{a^2 \arcsin(y/a)}{2} + \frac{y\sqrt{a^2 - y^2}}{2} \right) \Bigg|_0^b \\
&= \left(\frac{a^2 \arcsin(b/a)}{2} + \frac{b\sqrt{a^2 - b^2}}{2} \right) - 0 \\
&= \frac{a^2 \arcsin(b/a)}{2} + \frac{b\sqrt{a^2 - b^2}}{2}
\end{aligned}$$

A.4.7 Summary of Trigonometric Substitution

Here are three basic forms which are integrated by trig substitution summarised

Form	Substitute	Result
$\sqrt{a^2 - x^2}$	$x = a\cos\theta$ or $x = a\sin\theta$	$a\sin\theta$ or $a\cos\theta$
$\sqrt{a^2 + x^2}$	$x = a\tan\theta$	$a\sec\theta$
$\sqrt{x^2 - a^2}$	$x = a\sec\theta$	$a\tan\theta$

Ugly integrals expressed in such a form are candidates for these methods.

A.4.8 Partial Fractions

Partial fractions allow for integration of functions of the form

$$\frac{P(x)}{Q(x)}$$

Where $P(x)$ and $Q(x)$ are polynomials. Functions of this type are called *rational functions*. The method of partial fractions works by algebraically splitting $P(x)/Q(x)$ into pieces that are easier to integrate. Consider

$$\frac{4x-1}{(x-1)(x+2)} = \frac{A}{x-1} + \frac{B}{x+2}$$

We can solve for A by:

$$\begin{aligned} \frac{4x-1}{(x-1)(x+2)} &= \frac{A}{x-1} + \frac{B}{x+2} \\ \frac{4x-1}{x+2} &= A + \frac{B}{x+2}(x-1) \end{aligned}$$

And substituting $x = 1$ to get

$$\frac{4(3)-1}{1+2} = A$$

(next page)

Repeated factors
When faced with repeated factors in the denominator such as:

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

We require the addition of a second term for the repeated factor:

$$\frac{x^2+2}{(x-1)^2(x+2)} = \frac{A}{x-1} + \frac{B}{(x-1)^2} + \frac{C}{x+2}$$

The same applies for factors repeated more times. Intuitively, this is the same as:

$$\frac{7}{16} = \frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4}$$

Quadratic Factors
Consider

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

Notice the denominator cannot be split into any further factors (other than using imaginary numbers). In this case we split the fraction as follows:

$$\frac{x^2}{(x-1)(x^2+1)} = \frac{A}{x-1} + \frac{Bx+C}{x^2+1}$$

The point here is that the numerator corresponding to the quadratic factor is linear, not constant; generally it will be a polynomial with a degree one less than that of the denominator.

A.4.9 Integration by Parts

Integration of parts can be seen as the reverse of the product rule:

$$(uv)' = u'v + uv'$$

We derive the formula for integration by parts from the product rule:

$$\begin{aligned}
(uv)' &= u'v + uv' \\
uv' &= (uv)' - u'v \\
\int uv'dx &= \int (uv)'dx - \int u'v\,dx \\
\int uv'dx &= uv - \int u'v\,dx
\end{aligned}$$

In the case of definite integrals:

$$\int_a^b uv'dx = uv|_a^b - \int_a^b u'v\,dx$$

Example: $\int \ln x\,dx$
Where $u = \ln x$, $v = 1$, we get:

$$\int \underbrace{\ln x}_{uv'}\ dx = \underbrace{(\ln x)(x)}_{uv} - \int \underbrace{(\frac{1}{x})(x)}_{u'v}\ dx$$

and therefore

$$\int \ln x\,dx = x\ln x - x + c$$

(next page)

Reduction Formulas
Occasionally, integrals can be solved using *reduction formulas*, where we apply a rule to rewrite an integral in terms of another, simpler, integral. Here are some examples:

Example: $\int(\ln x)^n dx$
Integrating by parts:

$$\begin{aligned}\int(\ln x)^n dx &= x(\ln x)^n - n\int(\ln x)^{n-1}\frac{1}{x}\cdot x\,dx\\ &= x(\ln x)^n - n\int(\ln x)^{n-1}\,dx\end{aligned}$$

Therefore,

$$F_n(x) = x(\ln x)^n - nF_{n-1}(x) \quad \text{where } F_i = \int(\ln x)^i dx$$

Example: $\int x^n e^x dx$ Replacing e^x by $\sin x$ or $\cos x$ also leads to a similar derivation:

$$\int x^n e^x dx = x^n e^x - \int nx^{n-1}e^x dx$$

where we get

$$F_n(x) = x^n e^x - nF_{n-1}(x) \quad \text{where } F_i = \int x^i e^x dx$$

A.4.10 Arc length

Arc length can be calculated with a cumulative sum. Consider:

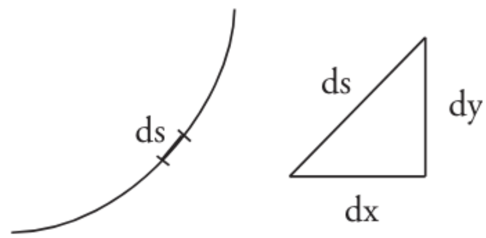


Figure: Straight line approximation of arc length

We can approximate

$$(\Delta s)^2 \approx (\Delta x)^2 + (\Delta y)^2$$

To get (as $\Delta s \to 0$)

$$ds = \sqrt{dx^2 + dy^2}$$

We can factor out dx to get

$$\frac{ds}{dx} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}$$

Therefore, if we have the x-coordinate range over our desired arc length:

$$\begin{aligned}\text{Arc Length} &= \text{Distance along the curve from } s_a \text{ to } s_b \\ &= \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx = \int ds \\ &= \int_a^b \sqrt{1 + f'(x)^2} dx\end{aligned}$$

Note that we are summing up changes in s, and not y. Thus we find arc length, not the area under the function.
(next page)

Example: Circular Arc
Consider finding the arc length of a circle (centered at origin and radius 1), on the interval $0 \leq x \leq a$:

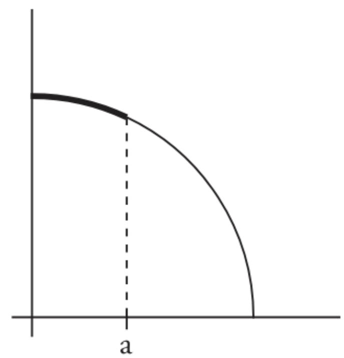


Figure: Arc length of circle over $0 \leq x \leq a$

Given our arc length formula, we need to find $f'(x)$:

$$\begin{aligned} y &= \sqrt{1-x^2} \\ y' &= \frac{-x}{\sqrt{1-x^2}} \\ \frac{ds}{dx} &= \sqrt{1+(y')^2} \\ &= \sqrt{1+\left(\frac{-x}{\sqrt{1-x^2}}\right)^2} \end{aligned}$$

We can simplify $1+\left(\frac{-x}{\sqrt{1-x^2}}\right)^2$:

$$\begin{aligned} \left(\frac{-x}{\sqrt{1-x^2}}\right)^2 &= 1+\frac{x^2}{1-x^2} \\ &= \frac{1}{1-x^2} \end{aligned}$$

to get

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

(next page)

Our arc length, α, is therefore given by

$$\begin{aligned}
\alpha &= \int_0^a \sqrt{\frac{1}{1-x^2}}dx \\
&= \sin^{-1} x\Big|_0^a \quad \text{(see A.1.20)} \\
\alpha &= \sin^{-1} a \quad \text{and} \quad \sin\alpha = a
\end{aligned}$$

Notice that when our radius is 1, **the angle in radians is equal to our arc length**

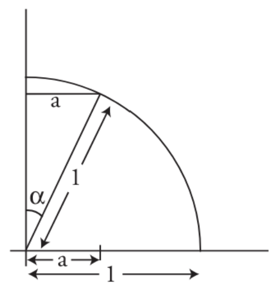


Figure: angle (in radians) = arc length = α

A.4.11 Parametric Equations

Parametric curves are essentially curves described by x and y being a function of a third variable t

$$
\begin{aligned}
x &= x(t) \\
y &= y(t)
\end{aligned}
$$

For instance,

$$
\begin{aligned}
x &= a\cos t \\
y &= a\sin t
\end{aligned}
$$

Describes a circle:

$$
\begin{aligned}
x^2 + y^2 &= (x(t))^2 + (y(t))^2 \\
&= a^2\cos^2 t + a^2\sin^2 t \\
x^2 + y^2 &= a^2
\end{aligned}
$$

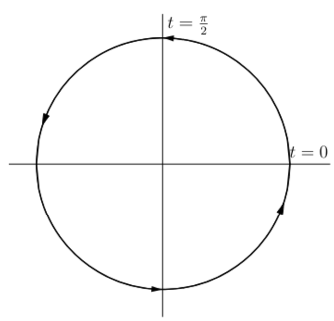


Figure: $(a\cos t, a\sin t)$

Example: Arc length Consider computing the arc length of the path of this trajectory. We have

$$
\begin{aligned}
ds &= \sqrt{dx^2 + dy^2} \\
\frac{ds}{dt} &= \frac{1}{dt}\sqrt{dx^2 + dy^2} \\
ds &= \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\,dt
\end{aligned}
$$

(next page)

since

$$\begin{aligned} x &= a\cos t, & \frac{dx}{dt} &= -a\sin t \\ y &= a\sin t, & \frac{dy}{dt} &= -a\cos t \end{aligned}$$

We have

$$\begin{aligned} ds &= \sqrt{(-a\sin t)^2 + (a\cos t)^2}dt \\ &= \sqrt{a^2}dt \\ &= a\,dt \end{aligned}$$

Should we define t as time, we can conclude that a point moves around the circle at speed $\frac{ds}{dt} = a$, constant speed.

A.4.12 Polar coordinates

Polar coordinates are another way of describing points in a plane:

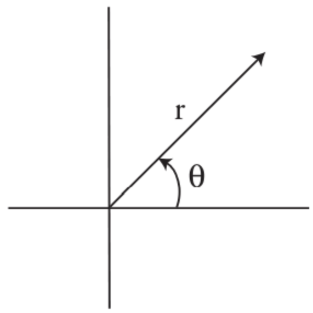


Figure: Polar coordinates describe a radius r and angle θ

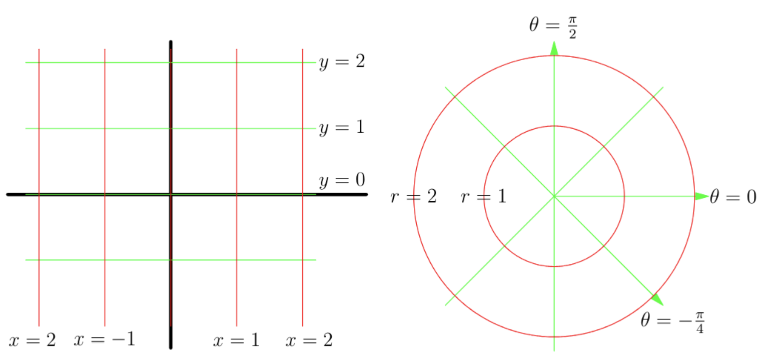
Unlike x, y coordinates (called the rectangular coordinate system), which are essentially coordinates on a grid, Polar coordinates are defined by an angle and a distance to theorigin Intuitively, therefore, conversion from polar coordinates to rectangular coordinates goes like

$$x = r\cos\theta, \quad y = r\sin\theta$$

also intuitively, conversion from rectangular coordinates to polar coordinates:

$$r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1}\left(\frac{y}{x}\right)$$

Note this statement has ambiguity—r could be $-\sqrt{x^2+y^2}$ and θ could be $\tan^{-1}\left(\frac{-y}{-x}\right)$. A diagramatic understanding of the coordinate is therefore also required. Understand that both coordinate systems describe the *same* space, just in different manners:



Figures: Rectangular (left) vs Polar (right) coordinate systems

(next page)

Example: Translation into Polar Coordinates
Consider translating $y = 1$ into polar coordinates:

$$\begin{aligned} y &= r \sin\theta \\ 1 &= r \sin\theta \\ r &= \frac{1}{\sin\theta} \end{aligned}$$

Illustrated:

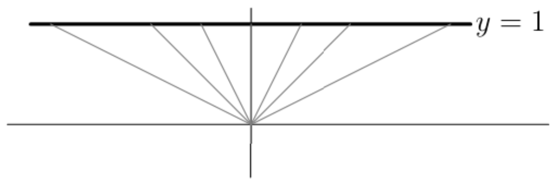


Figure: $r = \frac{1}{\sin\theta}$

Note here that as θ approaches 0 or $\pi/2$, r tends to infinity. This gives us our final answer:

$$r = \frac{1}{\sin\theta}, \quad 0 \le \theta \le \pi$$

A.4.13 Polar Coordinates and Area

We can directly calculate an area using polar coordinates. Similar to using Riemann sums, we divide the curve into 'slices':

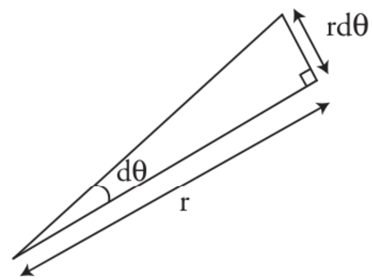


Figure: A slice—radius r and angle $d\theta$

where we can approximate each slice as a circular arc of arc length $rd\theta$. We define the incremental area of each slice as dA. To find dA, we use the idea that the proportion of arc length is equal to the proportion of total area covered:

$$\frac{dA}{\pi r^2} = \frac{rd\theta}{2\pi r}$$
$$dA = \frac{d\theta}{2\pi} \cdot \pi r^2$$
$$dA = \frac{1}{2} r^2 d\theta$$

Our final expression therefore looks like

$$A = \int_{\theta_1}^{\theta_2} \frac{1}{2} r^2 d\theta$$

Geometrically this looks like

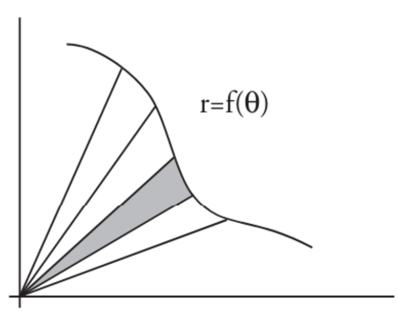


Figure: Cumulative sum of area in 'slices' from curve to origin

A.5 L'Hospital's Rule, Improper Integrals, Taylor Series

A.5.1 L'Hospital's Rule

L'Hospital's rule provides a systematic way of finding limits. Consider'the limit

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

L'Hospital's rule states that if $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$ or $\pm\infty$, and $g'(x) \neq 0$, then

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)}$$

Intuition:
Consider

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

where the limit is indeterminate, for instance $f(a) = g(a) = 0$. We can show

$$\begin{aligned}
\lim_{x\to a} \frac{f(x)}{g(x)} &= \lim_{x\to a} \frac{f(x)/(x-a)}{g(x)/(x-a)} \\
&= \frac{\lim_{x\to a} \frac{f(x)-f(a)}{x-a}}{\lim_{x\to a} \frac{g(x)-g(a)}{x-a}} \quad \text{(since } f(a) = g(a) = 0) \\
&= \frac{f'(a)}{g'(a)} \quad \text{(thus } g'(x) \neq 0)
\end{aligned}$$

Example:
Consider

$$\lim_{x\to 0} \frac{\sin 5x}{\sin 2x}$$

Since $a = 0$, $f(a) = g(a) = 0$, we apply L'Hospital's rule:

$$\begin{aligned}
\lim_{x\to 0} \frac{\sin 5x}{\sin 2x} &= \lim_{x\to 0} \frac{5\cos 5x}{2\cos 2x} \\
&= \frac{5\cos(0)}{2\cos(0)} \\
&= \frac{5}{2}
\end{aligned}$$

(next page)

Repeating L'Hospital's Rule

If

$$\lim_{x\to 0} \frac{f'(x)}{g'(x)}$$

is still indeterminate, L'Hospital's rule (if applicable) can be repeatedly used until a value is obtained. For instance, consider

$$\lim_{x\to 0} \frac{\cos x - 1}{x^2}$$

Since $f(a) = g(a) = 0$, we can apply L'Hospital's rule; notice however that the result is still indeterminate:

$$\lim_{x\to 0} \frac{\cos x - 1}{x^2} = \lim_{x\to 0} \frac{-\sin x}{2x}$$

Since $f'(a) = g'(a) = 0$, we can apply L'Hospital's rule again:

$$\begin{aligned} \lim_{x\to 0} \frac{\cos x - 1}{x^2} &= \lim_{x\to 0} \frac{-\sin x}{2x} \quad \text{(L'Hop)} \\ &= \lim_{x\to 0} \frac{-\cos x}{2} \quad \text{(L'Hop)} \\ &= \frac{-\cos 0}{2} \\ &= -\frac{1}{2} \end{aligned}$$

Note that we only know that the hypotheses of L'Hospital's rule hold after we compute the limit—the theorem states that L'Hospital's rule only works if the limit exists.

L'Hospital's rule extended:

Note that L'Hospital's rule also works for

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

Essentially it works for the $\frac{\infty}{\infty}$ case as well.

A.5.2 Examples of L'Hospital's Rule

Example: $\lim_{x\to 0^+} x \ln x$
Notice we are multiplying a number that's getting smaller by one that's getting larger and larger; the result therefore depends on the rates of growth of each component.

$$\lim_{x\to 0^+} \underbrace{x}_{\to 0} \underbrace{\ln x}_{\to -\infty}$$

We can rewrite the expression to apply L'Hospital's rule:

$$\begin{aligned}
\lim_{x\to 0^+} x \ln x &= \lim_{x\to 0^+} \frac{\ln x}{1/x} \\
&= \lim_{x\to 0^+} \frac{1/x}{-1/x^2} \quad \text{(L'Hop)} \\
&= \lim_{x\to 0^+} -x \\
&= 0
\end{aligned}$$

Intuitively this means x goes to 0 faster than $\ln x$ goes to $-\infty$, which may not have been apparent.

Example: $\lim_{x\to 0+} x^x$
We can rewrite:

$$x^x = e^{x \ln x}$$

to get

$$\begin{aligned}
\lim_{x\to 0+} x^x &= \lim_{x\to 0+} e^{x \ln x} \\
&= e^{\lim_{x\to 0+} x \ln x} \\
&= e^0 \\
&= 1
\end{aligned}$$

Whenever applying L'Hospital's rule, one has to verify that *they have an indeterminate form* for the rule to be valid.

A.5.3 L'Hospital's rule for rates of Growth and Decay

We can apply L'Hospital's rule to compare the rate at which functions change:

Rates of Growth
If $f(x) > 0$ and $g(x) > 0$ as x approaches infinty, then

$$f(x) << g(x) \text{ as } x \to \infty \quad \text{means} \quad \lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$$

For instance, since

$$\begin{aligned}\lim_{x\to\infty} \frac{\ln x}{x^2} &= \lim_{x\to\infty} \frac{1/x}{2x} \quad \text{(L'Hop)} \\ &= \lim_{x\to\infty} \frac{1}{2x^2} \\ &= 0\end{aligned}$$

we can conclude that $\ln x << x^2$ as $x \to \infty$. Other examples include: If $p > 0$ then

$$\ln x << x^p << e^x << e^{x^2} \text{ as } x \to \infty$$

Rates of Decay Similarly, we can compare rates at which functions tend to 0 as $x \to \infty$. If $p > 0$ then

$$\frac{1}{\ln x} >> \frac{1}{x^p} >> e^{-x} >> e^{-x^2} \text{ as } x \to \infty$$

A.5.4 Improper Integrals

Improper integrals are definite integrals where one (or both) of the boundaries is at infinity, or where the integrand has a vertical asymptote in the interval of integration. A possible improper integral of a function $f(x) > 0$ is

$$\int_a^\infty f(x)dx$$

which can also be written as a limit:

$$\lim_{N\to\infty}\int_a^N f(x)dx$$

The improper integral *converges* if this limit exists and *diverges* otherwise. Intutively, the geometric interpretation of the improper integral is the area under the curve stretching to infinity; if the integral converges the defined area under the curve is finite; otherwise it's infinite.

Example: $\int_1^\infty \frac{1}{x}dx$
We simply evaluate the integral (we let $N = \infty$ afterward):

$$\begin{aligned}\int_1^N \frac{1}{x}dx &= \ln x|_1^\infty \\ &= \ln(N) - \ln 1 \\ &= \ln(N) \to \infty \text{ as } N \to \infty\end{aligned}$$

This isn't quite intuitive, but we conclude that

$$\int_1^N \frac{1}{x}dx \text{ diverges}$$

It turns out that although $1/x$ diverges, $1/x^2$ *converges*. This is not intuitively visible, necessitating methods to determine convergence.

A.5.5 Integral Comparison

During situations where we can't directly compute the area, we can use Integral/Limit Comparison to compare that function to another whose value we can compute. This will allow us to tell us whether the former function converges or diverges

Consider that if $f(x) \sim g(x)$ as $x \to \infty$, then $\int_a^\infty f(x)\,dx$ and $\int_a^\infty g(x)\,dx$ either both converge or both diverge. ($f(x) \sim g(x)$ as $x \to \infty$ means $\frac{f(x)}{g(x)} \to 1$ as $x \to \infty$)

If we are unable to directly compute the integral of $g(x)$ to determine its convergence, we can instead compute that of $f(x)$, which converges in the same manner as g, in order to circumvent the original problem.

Example: Use limit comparison to show $\int_1^\infty \frac{1}{(5x+2)^2} dx$ converges
We require a function $f(x)$ comparable to our given function $g(x) = \frac{1}{(5x+2)^2}$. One approach is to expand the denominator and discard its lower degree terms, giving us $g(x) = \frac{1}{25x^2}$. First we check that $f(x) \sim g(x)$:

$$\begin{aligned}
\lim_{x\to\infty} \frac{f(x)}{g(x)} &= \lim_{x\to\infty} \frac{\frac{1}{25x^2}}{\frac{1}{(5x+2)^2}} \\
&= \lim_{x\to\infty} \frac{25x^2+20x+4}{25x^2} \\
&= \lim_{x\to\infty} \left(1 + \frac{4}{5x} + \frac{4}{25x^2}\right) \\
\lim_{x\to\infty} \frac{f(x)}{g(x)} &= 1
\end{aligned}$$

Since $g(x)$ only converges if $f(x)$ does, we only need to determine the simple integral

$$\begin{aligned}
\int_1^\infty \frac{1}{25x^2} dx &= \frac{1}{25} \int_1^\infty x^{-2} dx \\
&= \frac{1}{25} \left[-x^{-1}\right]_1^\infty \\
\int_1^\infty \frac{1}{25x^2} dx &= \frac{1}{25}
\end{aligned}$$

We conclude that

$$\int_1^\infty \frac{1}{(5x+2)^2} dx$$

must converge because

$$\int_1^\infty \frac{1}{25x^2} dx$$

converges

A.5.6 Improper Integrals and Singularities

Here we consider integration near singular points. Integrals like these with discontinuities in the interval of integration are considered *indefinite integrals of the second type*. Examples include:

$$\int_0^1 \frac{dx}{\sqrt{x}}, \quad \int_0^1 \frac{dx}{x} \quad , \int_0^1 \frac{dx}{x^2}$$

These integrals turn out to be fairly straightforward to calculate:

$$\begin{aligned}\int_0^1 \frac{dx}{\sqrt{x}} &= \int_0^1 x^{-1/2} dx \\ &= \frac{1}{1/2} x^{1/2}\Big|_0^1 \\ &= 2\end{aligned}$$

$$\begin{aligned}\int_0^1 \frac{dx}{x} &= \ln x|_0^1 \\ &= \ln 1 - \ln 0 \quad \text{(diverges)}\end{aligned}$$

$$\begin{aligned}\int_0^1 \frac{dx}{x^2} &= -x^{-1}|_0^1 \\ &= -1 + \left(\frac{1}{0}\right) \quad \text{(diverges)}\end{aligned}$$

(Notice how although the three functions intuitively have a similar shape, one of them converges while the others diverge)

Note that we have to be careful near singularities—integrating over the singularity will lead to a nonsensical answer:

$$\int_{-1}^1 \frac{dx}{x^2} = -2 \quad \text{despite } \frac{1}{x^2} > 0$$

Fact is that the area here is infinite, not -2
(next page)

Improper Integrals of the Second Kind
Similar to other improper integrals, we can express integrals near singularities as limits; consider $f(x) = \frac{1}{x}$, which has a singularity at 0:

$$\int_0^1 f(x)\,dx = \lim_{a \to 0^+} \int_a^1 f(x)\,dx$$

As before, we say the integral *converges* if this limit exists and *diverges* if not:

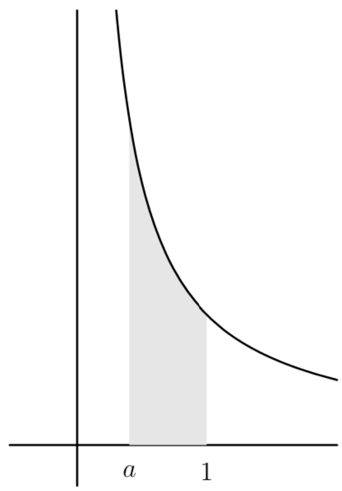


Figure: Area under the graph of $y = \frac{1}{x}$

In general, regarding the singularity near 0 for $\frac{1}{x^p}$:

$$\begin{aligned}
\int_0^1 \frac{1}{x^p} &= \left.\frac{x^{-p+1}}{-p+1}\right|_0^1 \quad \text{(for } p \neq 1\text{)} \\
&= \frac{1^{-p+1}}{-p+1} - \frac{0^{-p+1}}{-p+1} \\
&= \begin{cases} \frac{1}{-p+1} & \text{if } p < 1 \\ \text{diverges} & \text{if } p \geq 1 \end{cases}
\end{aligned}$$

(next page)

Some intuition
Despite being of a similar shape, the limits of $\int \frac{1}{x^p}$ differ depending on p. Here we propose an intuitive understanding; calculating limits, we find

$$\frac{1}{x^{-1/2}} << \underbrace{\frac{1}{x} << \frac{1}{x^2}}_{\text{diverges}} \quad \text{as } x \to 0^+$$
$$\underbrace{\frac{1}{x^{-1/2}} >> \frac{1}{x}}_{\text{diverges}} >> \frac{1}{x^2} \quad \text{as } x \to \infty$$

The idea is that $\frac{1}{x}$ is symmetric to itself by a reflection across the line $y = x$, and the areas under the curve for $x \geq 1$ and $0 \leq x \leq 1$ are both ∞.

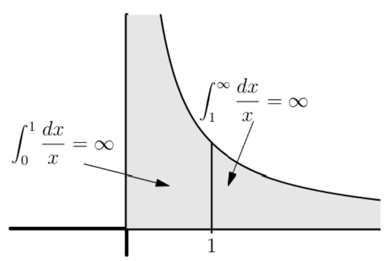
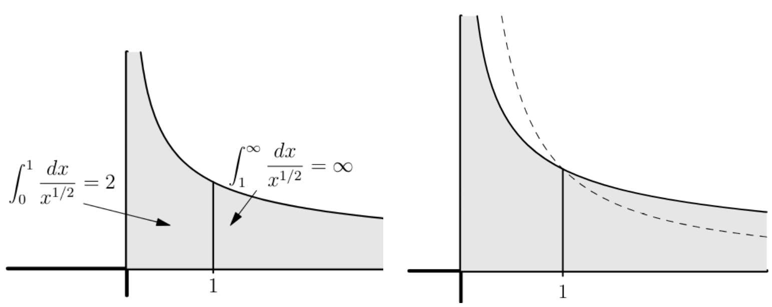


Figure: Area under the graph of $y = \frac{1}{x}$ is infinite in both directions

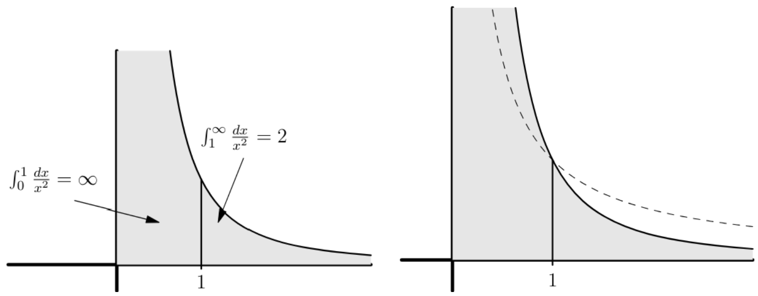
Now consider the $\frac{1}{x^{1/2}}$; it isn't symmetric about $y = x$, and one 'tail' larger than the other:



Left figure: Area under the graph of $y = \frac{1}{x^{1/2}}$ is infinite in only one direction
Right figure: $\frac{1}{x}$ superimposed onto $\frac{1}{x^{1/2}}$

Notice the difference when $\frac{1}{x}$ is superimposed onto $\frac{1}{x^{1/2}}$; comparing the 'tails' of the functions allows for a geometric sense of the difference between convergent and divergent integrals.
(next page)

Comparing the tails of $\frac{1}{x^2}$ and $\frac{1}{x}$, we see that the divergent nature of the 'tails' can also be seen here:



Left figure: Area under the graph of $y = \frac{1}{x^2}$ is infinite in only one direction
Right figure: $\frac{1}{x}$ superimposed onto $\frac{1}{x^2}$

A.6 Limits and Continuity

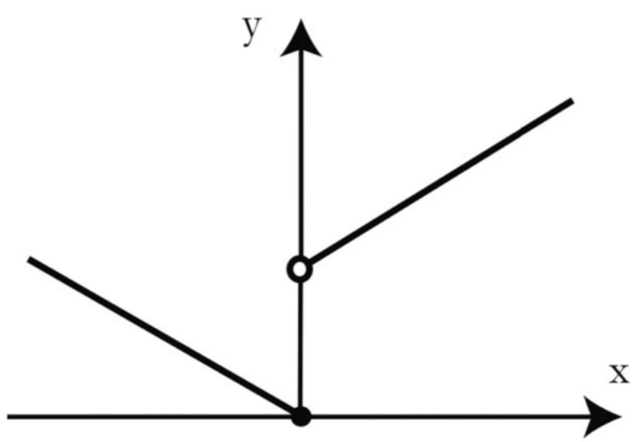
A.6.1 Limits

Limits can be further subdivided into right and left-handed limits. The limit:

$$\lim_{x \to x_0^+} f(x)$$

Is called a *right-hand limit*, meaning values to the right of x_0 on the number line should be used to compute the limit. For instance consider $f(x)$:

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



In this case the right-hand limit $\lim_{x \to 0^+} f(x)$ is 1. Similarly, the *left-hand limit*:

$$\lim_{x \to x_0^-} f(x)$$

is found when values to the *left* of x_0 on the number line are used to compute the limit. In this case, $\lim_{x \to 0^-} f(x) = 0$.

A.6.2 Continuity and Differentiability

Definition: A function is *continuous* at x_0 if $\lim_{x \to x_0} f(x) = f(x_0)$.
More specifically, a function continuous at x_0 has the properties:

- $\lim_{x \to 0^+} f(x) = \lim_{x \to 0^-} f(x)$; and both of these one sided limits exist.
- $f(x_0)$ is defined.
- $\lim_{x \to 0^+} f(x) = \lim_{x \to 0^-} f(x) = f(x_0)$.

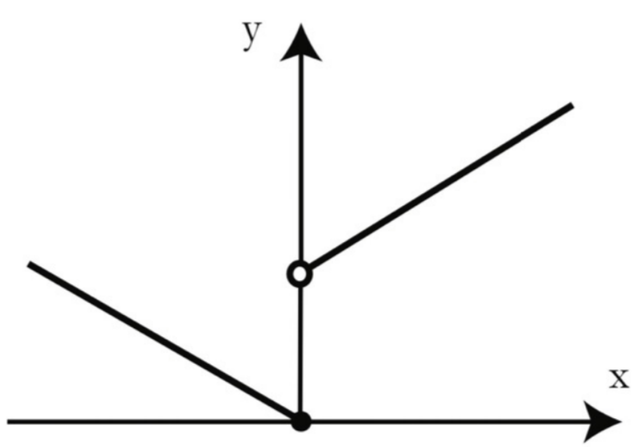
Remember that when calculating $\lim_{x \to x_0} f(x)$ one never allows x to equal x_0. Note that $\lim_{x \to x_0} f(x)$ is calculated differently and independently of $f(x_0)$; if one isn't careful to make this distinction, this definition has no meaning.

The same concept applies to **differentiability**, where a function is *differentiable* at x_0 if $\lim_{x \to x_0} f'(x) = f'(x_0)$, and:

- $\lim_{x \to 0^+} f'(x) = \lim_{x \to 0^-} f'(x)$; and both of these one sided limits exist.
- $f'(x_0)$ is defined.
- $\lim_{x \to 0^+} f'(x) = \lim_{x \to 0^-} f'(x) = f'(x_0)$.

A.6.3 Discontinuity

A non-continuous function implies the presence of a *discontinuity*. Here are a few types of discontinuities:
A **Jump Discontinuity** occurs when the right and left-hand limits exist but are not equal, the example in A.6.1 is an instance of this:



Although $\lim_{x \to 0^+} f(x)$ and $\lim_{x \to 0^-} f(x)$ exist, they are **not equal**; therefore this function is discontinuous.

A **Removable Discontinuity** occurs when the right and left-hand limits are equal but either the function is not defined or not equal to these limits:

$$\lim_{x \to 0^+} f(x) = \lim_{x \to 0^-} f(x) \neq f(x_0)$$

Figure: A removable discontinuity, continuous everywhere except one point.

An example of this is...

In an **Infinite discontinuity**, the left and right-hand limits are infinite, either limit may be negative or positive:

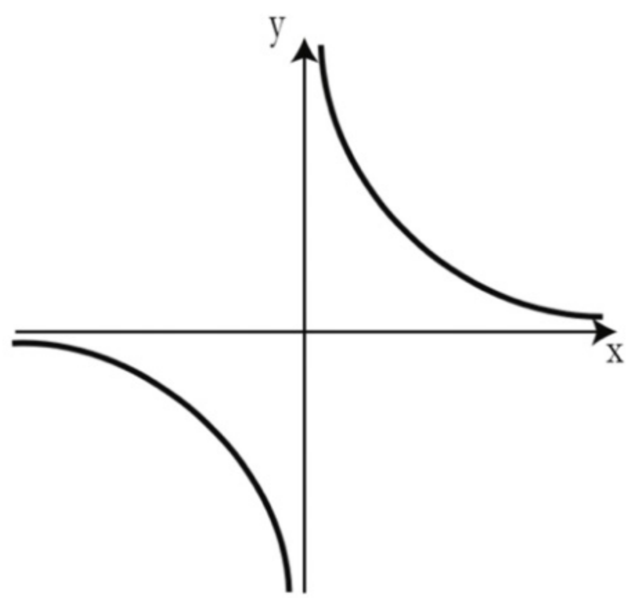


Figure: An infinite discontinuity: $\frac{1}{x}$

In this case, $\lim_{x \to 0^+} f(x) = \infty$, while $\lim_{x \to 0^-} f(x) = -\infty$ note that a limit of ∞ is different from saying a limit doesn't exist.

A.6.4 Differentiable implies Continuous

Theorem: If f is differentiable at x_0, then f is continuous at x_0.
Proof: We need to show

$$\lim_{x \to x_0} f(x) = f(x_0)$$
$$\lim_{x \to x_0} f(x) - f(x_0) = 0$$

We can prove this through:

$$\lim_{x \to x_0} f(x) - f(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}(x - x_0)$$

Notice that when $x \to x_0$, we can say

$$\lim_{x \to x_0} f(x) = \lim_{\Delta x \to 0} f(x_0 + \Delta x)$$

and that

$$\lim_{x \to x_0} x = \lim_{\Delta x \to 0} (x_0 + \Delta x)$$

going back to the original equation:

$$\begin{aligned}\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}(x - x_0) &= \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{(x_0 + \Delta x) - x_0}((x_0 + \Delta x) - x_0) \\ &= \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}(\Delta x)\end{aligned}$$

Now assuming that f is **differentiable** at x_0:

$$\begin{aligned}\lim_{x \to x_0} f(x) - f(x_0) &= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}(x - x_0) \\ &= \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}(\Delta x) \\ &= f'(x) \cdot 0 \\ &= 0\end{aligned}$$

Notice that our assumption that f was differentiable at x_0 leads to $\lim_{x \to x_0} f(x) - f(x_0) = 0$, indicating that differentiability implies continuity. Also note that dividing by $\lim_{x \to x_0}(x - x_0)$ is not the same as division by 0 since $\lim_{x \to x_0}(x - x_0)$ never truly reaches 0.

(Note that one might intuively understand that $\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x)$ in line 6. I've changed the limit to produce the difference quotient for ease of understanding.)

A.7 Infinite Series

A.7.1 Notation

We define a *partial sum* to be

$$S_N = \sum_{n=0}^{N} a_n$$

Similar to indefinite integrals, we can define

$$S = \sum_{n=0}^{\infty} a_n = \lim_{N\to\infty} S_N$$

Where if the limit exists the series *converges*, and otherwise *diverges*.

A.7.2 Integral Comparison for Series

Evaluating the limits of series is tends to be difficult. Integral comparison allows us to understand series limits in terms of integrals:

Theorem If $f(x)$ is decreasing and $f(x) > 0$ on the interval $[1, \infty]$, then the sum $\sum_1^\infty f(n)$ and the integral $\int_1^\infty f(x)\,dx$ diverge and converge together and

$$\sum_1^\infty f(n) - \int_1^\infty f(x)\,dx < f(1)$$

Example: Do the series $\sum_{n=2}^\infty \frac{1}{n \ln n}$ and $\sum_{n=2}^\infty \frac{1}{n(\ln n)^2}$ converge or diverge?

First we should check that the functions of interest are decreasing and positive on the specified interval. Then with the Integral comparison test we can say that since

$$\int_2^\infty \frac{1}{x \ln x} dx = \infty \quad \text{diverges}$$
$$\sum_{n=2}^\infty \frac{1}{n \ln n} \quad \text{diverges}$$

and that since

$$\int_2^\infty \frac{1}{x(\ln x)^2} dx = \ln 2 \quad \text{converges}$$
$$\sum_{n=2}^\infty \frac{1}{n(\ln n)^2} \quad \text{converges}$$

(next page)

Intuition for Integral comparison: Harmonic Series
Consider comparing:

$$\sum_{n=1}^{\infty} \frac{1}{n} \quad \text{with} \quad \int_1^{\infty} \frac{dx}{x}$$

Notice that the summation is just an upper Riemann sum of the integral with $\Delta x = 1$:

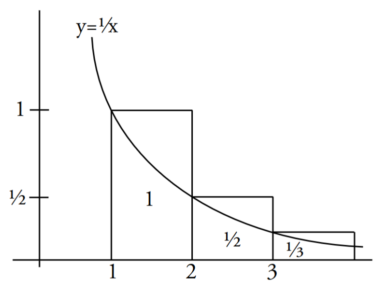


Figure: The summation *is* an upper Riemann sum of the function

Intuitively we can see that the upper Riemann sum is greater than the integral:

$$\int_1^N \frac{dx}{x} < 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{N-1}$$

(There are $N - 1$ rectangles because the distance between 1 and N is $N - 1$.)
If $S_N = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{N}$, we can further say

$$\int_1^N \frac{dx}{x} < 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{N-1} < S_N$$

This allows us to conclusively prove that the series diverges:

$$\int_1^N \frac{dx}{x} = \ln N < S_N$$

Since $\lim_{N\to\infty} \ln N = \infty$.
(next page)

Since for $f(x) = \frac{1}{x}$, $\ln N < S_N$ and $\ln N$ diverges,

$$\lim_{N\to\infty} S_N = \sum_{n=1}^{\infty} \frac{1}{n} \quad \text{diverges}$$

Now consider the lower Riemann sum:

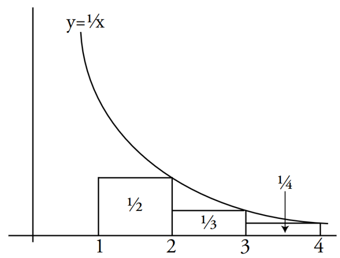


Figure: Lower Riemann sum of $\frac{1}{x}$

Intuitively we see that

$$\begin{aligned}
\int_1^N \frac{dx}{x} &> \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{N} = S_N - f(1) \\
\ln N &> S_N - f(1) \\
\ln N + f(1) &> S_N
\end{aligned}$$

Combining this with the result from the upper Riemann sum we can conclude

$$\begin{gathered}
\int_1^N \frac{dx}{x} = \ln N < S_N < \ln N + f(1) = \int_1^N \frac{dx}{x} + f(1) \\
\int_1^N \frac{dx}{x} < \sum_{n=1}^{N} \frac{1}{n} < \int_1^N \frac{dx}{x} + f(1) \\
\left| \sum_{n=1}^{N} \frac{1}{n} - \int_1^N \frac{dx}{x} \right| < f(1)
\end{gathered}$$

A.7.3 Limit Comparison for Series

Similar to with integrals, Limit comparison can be used with two series to determine convergence properties:

Theorem: If $f(n) \sim g(n)$ (i.e if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$) and $g(n) > 0$ for all n, then $\sum_{n=1}^{\infty} f(n)$ and $\sum_{n=1}^{\infty} g(n)$ converge or diverge together.

Example: Does $\sum_{n=0}^{\infty} \frac{5n+2}{n^3+1}$ converge or diverge?
We can compare

$$\sum_{n=0}^{\infty} \frac{5n+2}{n^3+1} \quad \text{with} \quad \sum_{n=1}^{\infty} \frac{5n}{n^3} = \sum_{n=1}^{\infty} \frac{5}{n^2}$$

We should check the $f(n) \sim g(n)$:

$$\begin{aligned} \lim_{n\to\infty} \frac{f(n)}{g(n)} &= \lim_{n\to\infty} \frac{\frac{5n+2}{n^3+1}}{\frac{5}{n^2}} \\ &= \lim_{n\to\infty} \frac{5n^3+2n^2}{5n^3+5} = 1 \end{aligned}$$

Thus since $\sum_{n=1}^{\infty} \frac{5}{n^2}$ converges, $\sum_{n=0}^{\infty} \frac{5n+2}{n^3+1}$ also converges.

A.7.4 Ratio Test for convergence of series

A ratio text can be used to determine the convergence properties of a series. Say we have a series

$$\sum a_n \quad \text{where } a_n > 0$$

Now consider the limit:

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = L$$

We can say that if

$$L < 1, \quad \sum a_n \text{ converges}$$
$$L > 1, \quad \sum a_n \text{ diverges}$$

(For $L = 1$ different cases can occur and we can't make any deductions)

Intuition
Consider the case where

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = L < 1$$

We can say for some large $n > n_0$ (n_0 just defines some threshold for n)

$$\frac{a_{n+1}}{a_n} < r = \frac{r^{n+1}}{r^n}$$

r is a ratio with value $L < r < 1$. Since the limit of $\frac{a_{n+1}}{a_n}$ tends to L, we can say that for some high enough n eventually this criteria is met. With that we get

$$\frac{a_{n+1}}{a_n} < \frac{r^{n+1}}{r^n}$$
$$\frac{a_{n+1}}{r^{n+1}} < \frac{a_n}{r^n}$$

Note the meaning of this: the ratio between a_n and r^n is decreasing. Therefore for large enough n:

$$a_n < kr^n$$

(remember this isn't necessarily true all the way throughout the series), but now we can say

$$\sum_{n=n_0} a_n < k \sum_{n=n_0} r^n$$

since we defined r to be $L < r < 1$, $k \sum_{n=n_0} r^n$ converges, and therefore so does $\sum_{n=n_0} a_n$. This means that

$$\sum a_n \quad \text{converges}$$

(next page)

Example: Does $\sum \frac{4^n}{n3^n}$ converge or diverge?

$$\begin{aligned}
\lim_{n\to\infty} \frac{a^{n+1}}{a_n} &= \lim_{n\to\infty} \frac{4^{n+1}}{(n+1)3^{n+1}} \cdot \frac{n3^n}{4^n} \\
&= \lim_{n\to\infty} \frac{4n}{3(n+1)} \\
&= \frac{4}{3} > 1 \quad \text{(diverges)}
\end{aligned}$$

Example: Does $\sum \frac{n^10}{10^n}$ converge or diverge?

$$\begin{aligned}
\lim_{n\to\infty} \frac{(n+1)^{10}}{10^{n+1}} \cdot \frac{n^{10}}{10^n} &= \lim_{n\to\infty} \frac{1}{10} \cdot \frac{(n+1)^{10}}{n^{10}} \\
&= \frac{1}{10} < 1 \quad \text{(converges)}
\end{aligned}$$

A.7.5 Power Series

General Power Series and Radius of Convergence:
In general, a power series is of the form

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots = \sum_{n=0}^{\infty} a_n x^n$$

A power series converges when $-R < x < R$, where R is called the *radius of convergence*; the value of R depends on the coefficients a_i. When $|x| > R$, the series diverges; correspondingly when $|x| < R$, the series converges.

When $|x| < R$, the values of each term in the series will tend to 0 exponentially fast. While in the $|x| > R$ case the values won't tend to 0 at all.

Example: Consider the power series with the sum:

$$1 + x + x^2 + x^3 + \ldots = \frac{1}{1-x} \quad (|x| < 1)$$

We prove the sum here; denoting S to be the sum:

$$\begin{aligned} 1 + x + x^2 + x^3 + \ldots &= S \\ x + x^2 + x^3 + x^4 + \ldots &= Sx \\ 1 + 0 + 0 + \ldots &= S - Sx \end{aligned}$$

(Note that Sx contains one more higher degree x than S, but since high degree terms tend to 0 (since S converges as it tends to a constant), that higher degree term tends to 0.)

$$\begin{aligned} 1 &= S - Sx \\ S &= \frac{1}{1-x} \end{aligned}$$

If it wasn't obvious already, S has to converge in order for it to exist; this leads to the requirement for $|x| < 1$—the radius of convergence in this case is 1.
(next page)

Ratio Test to find Radius of Convergence
One can apply a ratio test to find the radius of convergence. For a series $\sum_{n=n_0}^{\infty} c_n x^n$ we can apply the ratio test to find

$$L = \lim_{n\to\infty} \left| \frac{c_{n+1}x^{n+1}}{c_n x^n} \right|$$

(Note the ratio test still intuitively makes sense even when taking the absolute value of the limit.) Point here is that we are able to find the value of x for which L crosses 1—meaning the point where the series stops converging and starts diverging, giving us the radius of convergence.

Example: Radius of Convergence of $\sum_{n=1}^{\infty} \frac{x^n}{n}$
In this case

$$\begin{aligned}
\frac{c_{n+1}x^{n+1}}{c_n x^n} &= \frac{x^{n+1}/(n+1)}{x^n/n} \\
&= x \cdot \frac{n}{n+1}
\end{aligned}$$

and so our ratio test gives us

$$\begin{aligned}
L &= \lim_{n\to\infty} \left| \frac{c_{n+1}x^{n+1}}{c_n x^n} \right| \\
&= \lim_{n\to\infty} \left| x \cdot \frac{n}{n+1} \right| \\
&= \lim_{n\to\infty} \left| x \cdot \left(1 - \frac{1}{n+1}\right) \right| \\
L &= |x|
\end{aligned}$$

When $|x| < 1, L < 1$ and the series will converge. When $|x| > 1, L > 1$ and the series diverges—the radius of convergence is 1.

A.7.6 Taylor's Series

Taylor's formula
The formula is given as

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$$

Intuition
Consider

$$\begin{aligned}
f(x) &= a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots \\
f'(x) &= a_1 + 2a_2 x + 3a_3 x^2 + \dots \\
f''(x) &= (1 \cdot 2)a_2 + (2 \cdot 3)a_3 x + (3 \cdot 4)a_4 x^2 + \dots \\
f^{(3)}(x) &= (1 \cdot 2 \cdot 3)a_3 + (2 \cdot 3 \cdot 4)a_4 x + \dots
\end{aligned}$$

Notice that by substituting in $x = 0$,

$$f^{(n)}(0) = (n!)a_n$$

Thus

$$a_n = \frac{f^{(n)}(0)}{n!}$$

Examples
Using this procedure we can evaluate:

$$\begin{aligned}
e^x &= \sum_{n=0}^{\infty} \frac{1}{n!} x^n \\
\sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots \\
\cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots
\end{aligned}$$

More intuition:
Note that a more general form of the Taylor expansion can also be considered:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!} (x - c)^n$$

(expansion at $x = c$ instead of 0) A finite Taylor series expansion is simply a n-degree approximation at a specific point (think of it as being similar to Fourier series!). Note that although an 'infinite' expansion should equal the function of interest, it is the *limit* of the series that approaches the function—we won't ever have a 'closed form expression' using these expansions.

A.7.7 When a function is equivalent to its Taylor series

We address the question of when a function is equal to its Taylor Series. Recall that a series can be expressed as limits:

$$\sum_{n=0}^{\infty} a_n = \lim_{N\to\infty} \sum_{n=0}^{N} a_n$$

In the context of Taylor series $T_f(x)$ for a function f, the Taylor series can be expressed as the limit of a limit of Taylor Polynomials $P_N(x)$ as $N \to \infty$:

$$T_f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!}(x-c)^n = \lim_{N\to\infty} \sum_{n=0}^{N} \frac{f^{(n)}(x)}{n!}(x-c)^n = \lim_{N\to\infty} P_N(x)$$

Where $P_N(x)$ is the N-th degree Taylor polynomial to $f(x)$ at $x = c$. Notice $f(x)$ will equal $T_f(x)$ provided

$$f(x) - \lim_{N\to\infty} P_N(x) = \lim_{N\to\infty} (f(x) - P_N(x)) = 0$$

We can denote $f(x) - P_N(x)$ to be the 'remainder' $R_N(x)$. A function therefore equals its Taylor series when

$$\lim_{N\to\infty} R_N(x) = 0$$

To proceed we use Taylor's inequality.

Taylor's inequality. *If $f(x)$ is n-times differentiable and $|f^{(n)}(x)| \leq M$ for all $x \in [c, c+d]$, then*

$$|f(x) - P_{n-1}(x)| \leq \frac{M}{n!}(x-c)^n$$

Intuition: (base case $n = 1$) If $f^{(1)} \leq M$ for $x \in [c, c+d]$,

$$\begin{aligned}
\int_c^x f^{(1)}(t)dt &\leq \int_c^x M\,dt \\
f(x) - f(c) &\leq M(x-c) \\
f(x) &\leq f(c) + M(x-c) \\
&= P_0(x) + \frac{M}{1!}(x-c)^1
\end{aligned}$$

(proof can be carried out by induction)
(next page)

Returning to the original problem, taylor's inequality gives us

If $f(x)$ is $(n+1)$-times differentiable and $|f^{(n+1)}(x)| \le M$ for all $x \in [c-d, c+d]$, then

$$|f(x) - P_N(x)| = |R_N(x)| \le \frac{M}{(N+1)!}|x-c|^{N+1}$$

For all $x \in [c-d, c+d]$.

Show thet e^x converges to its Taylor Series at 0 for all real x:
We first find some M where $|f^{(n+1)}(x)| \le M$ for all $x \in [c-d, c+d]$; knowing that the derivative of e^x is itself, an increasing function, and that we are interested in convergence to the taylor series near 0, a suitable bound would be:

$$|f^{(N+1)}(x)| \le e^d \quad \text{for all } x \text{ in } [-d, d]$$

Using e^d in our inequality:

$$\begin{aligned}
\lim_{N\to\infty} |R_N(x)| &\le \lim_{N\to\infty} \frac{e^d}{(N+1)!}|x|^{N+1} \\
&= \lim_{N\to\infty} e^d \frac{|x|^{N+1}}{(N+1)!} \\
&= 0
\end{aligned}$$

A.8 Differential Equations

A.8.1 Introduction to Ordinary Differential Equations (ODEs)

Here we introduce intuition for Ordinary Differential Equations (ODEs) and introductory solving methods.

The simplest type of differential equation looks like:

$$\frac{dy}{dx} = f(x)$$

which can be solved by the antiderivative $y = \int f(x)\,dx$.

Intuition
Now we consider a more interesting example:

$$\frac{dy}{dx} + xy = 0$$

This equation can be solved by *separation of variables*:

$$\begin{aligned}\frac{dy}{dx} + xy = 0\\ \frac{dy}{dx} &= -xy\\ \frac{dy}{y} &= -x\,dx\end{aligned}$$

Since the problem is now set up in terms of differentials rather than ratios of differentials, we can integrate both sides.

$$\begin{aligned}\int \frac{dy}{y} &= -\int x\,dx\\ \ln y + c_1 &= -\frac{x^2}{2} + c_2 \quad \text{(assume } y > 0)\end{aligned}$$

We can combine the constants and simplify:

$$\begin{aligned}\ln y &= -\frac{x^2}{2} + c\\ e^{\ln y} &= e^{-x^2/2+c}\\ y &= e^c e^{-x^2/2}\\ y &= Ae^{-x^2/2}, \quad \text{(where } A = e^c)\end{aligned}$$

(The more apt $\ln|y|$ simplifies to $\pm Ae^{-x^2/2}$, which doesn't matter since A is some unspecific constant)
(next page)

It turns out that our solution,

$$y = Ae^{-x^2/2}, \quad \text{(where } A = e^c\text{)}$$

Works for any constant multiple A. We can check this solution:

$$\begin{aligned}
y &= ae^{-x^2/2} \\
\frac{dy}{dx} &= \frac{d}{dx} ae^{-x^2/2} \\
&= a \cdot (-x)e^{-x^2/2} \\
&= -x \cdot ae^{-x^2/2} \\
\frac{dy}{dx} &= -xy
\end{aligned}$$

A is determined by an initial condition; for instance if $y(0) = 1$, $A = 1$.

A.8.2 Separation of Variables

Here we describe a rudimentary method for solving some differential equations—Separation of Variables.

In general, this method applies to differential equations of the form

$$\frac{dy}{dx} = f(x)g(y)$$

Where we then *separate* the variables and integrate:

$$\begin{aligned}
\frac{dy}{dx} &= f(x)g(y) \\
\frac{dy}{g(y)} &= f(x)\,dx \\
h(y)\,dy &= f(x)\,dx \quad \text{where } h(y) = \frac{1}{g(y)} \\
\int h(y)\,dy &= \int f(x)\,dx
\end{aligned}$$

Antidifferentiating both sides:

$$H(y) = \int h(y)\,dy; \quad F(x) = \int f(x)\,dx$$

we now have

$$\begin{aligned}
H(y) + c_1 &= F(x) + c_2 \\
H(y) &= F(x) + c
\end{aligned}$$

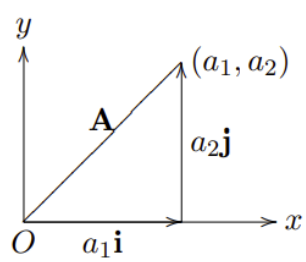
Appendix B

Multivariable Calculus/Calculus III

B.1 Vectors and Matrices

B.1.1 Notation, Terminology

For reference regarding notation for vectors and matrices:



We regard $\mathbf{A}$ as an *origin vector*, written as

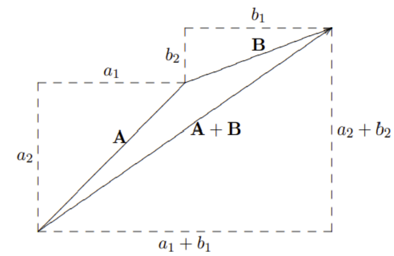
$$\mathbf{A} = \langle a_1, a_2 \rangle$$

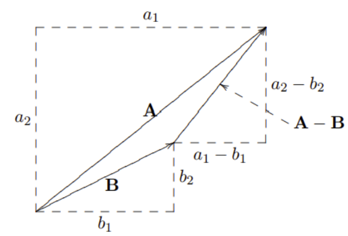
Vectors $\mathbf{i}$ and $\mathbf{j}$ denote *unit vectors* (vectors of magnitude 1) $\mathbf{i} = \langle 1, 0 \rangle$, $\mathbf{j} = \langle 0, 1 \rangle$. As such,

$$\langle a_1, a_2 \rangle = a_1\mathbf{i} + a_2\mathbf{j}$$

(next page)

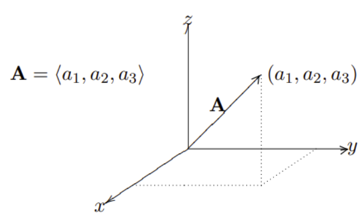
Vector algebra
As illustrated,

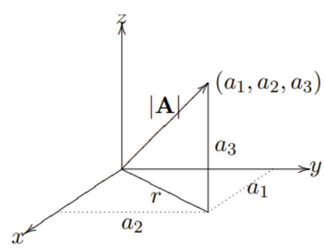




- Addition: $\mathbf{A}+\mathbf{B} = (a_1+b_1)\mathbf{i} + (a_2+b_2)\mathbf{j}$;
  $\langle a_1, a_2\rangle + \langle b_1, b_2\rangle = \langle a_1+b_1, a_2+b_2\rangle$
- Subtraction: $\mathbf{A}-\mathbf{B} = (a_1-b_1)\mathbf{i} + (a_2-b_2)\mathbf{j}$;
  $\langle a_1, a_2\rangle - \langle b_1, b_2\rangle = \langle a_1-b_1, a_2-b_2\rangle$
- Magnitude: $|\mathbf{A}| = \sqrt{a_1^2+a_2^2}$ (Pythagorean theorem)

In three dimensions:





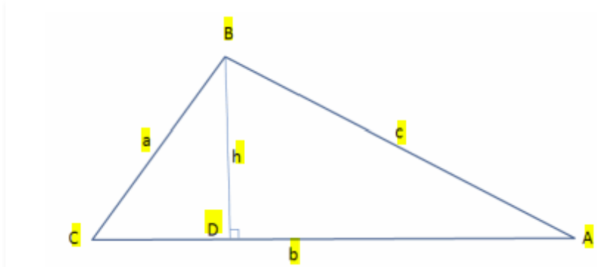
Similarly, we denote $\mathbf{A} = \langle a_1, a_2, a_3\rangle$, $\mathbf{i} = \langle 1,0,0\rangle$, $\mathbf{j} = \langle 0,1,0\rangle$, $\mathbf{k} = \langle 0,0,1\rangle$, where

$$\langle a_1, a_2, a_3\rangle = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$$

Magnitude in three dimensions also follows from the Pythagorean theorem:

$$|a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}| = |\langle a_1, a_2, a_3\rangle| = \sqrt{a_1^2+a_2^2+a_3^2}$$

B.1.2 Proof for Law of cosines



We show

$$c^2 = a^2 + b^2 - 2ab\cos C$$

Consider

$$\begin{aligned} CD &= a\cos C \\ DA &= b - a\cos C \quad (1) \\ BD &= a\sin C \quad (2) \end{aligned}$$

Using the pythagorean theorem:

$$\begin{aligned} c^2 &= DA^2 + BD^2 \\ &= (b - a\cos C)^2 + (a\sin C)^2 \\ &= b^2 - 2ab\cos C + a^2(\sin^2 C + \cos^2 C) \\ c^2 &= a^2 + b^2 - 2ab\cos C \end{aligned}$$

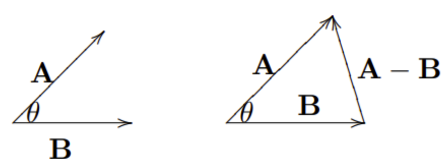
B.1.3 Dot product

The dot product has Algebraic and Geometric perspectives:
Algebraic view (for 2D vectors)
If $\mathbf{A} = \langle a_1, a_2 \rangle$ and $\mathbf{B} = \langle b_1, b_2 \rangle$ then

$$\mathbf{A} \cdot \mathbf{B} = a_1 b_1 + a_2 b_2$$

Geometric view
Given $\mathbf{A}, \mathbf{B}$ with angle θ between them:

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos\theta$$



Intuition
We can use the law of cosines to show that the two views are the same:

$$|\mathbf{A} - \mathbf{B}|^{\mathbf{2}} = |\mathbf{A}|^{\mathbf{2}} + |\mathbf{B}|^{\mathbf{2}} - 2|\mathbf{A}|\mathbf{B}| \cos\theta$$

given $\mathbf{A} = \langle a_1, a_2 \rangle$ and $\mathbf{B} = \langle b_1, b_2 \rangle$,

$$\begin{aligned}
2|\mathbf{A}|\mathbf{B}| \cos\theta &= |\mathbf{A}|^{\mathbf{2}} + |\mathbf{B}|^{\mathbf{2}} - |\mathbf{A} - \mathbf{B}|^{\mathbf{2}} \\
&= \left(\sqrt{a_1^2 + a_2^2}\right)^2 + \left(\sqrt{b_1^2 + b_2^2}\right)^2 - \left(\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}\right)^2 \\
&= a_1^2 + a_2^2 + b_1^2 + b_2^2 - [(a_1 - b_1)^2 + (a_2 - b_2)^2] \\
2|\mathbf{A}|\mathbf{B}| \cos\theta &= 2a_1 b_1 + 2a_2 b_2
\end{aligned}$$

this simplifies to

$$\begin{aligned}
a_1 b_1 + a_2 b_2 &= |\mathbf{A}|\mathbf{B}| \cos\theta \\
\mathbf{A} \cdot \mathbf{B} &= |\mathbf{A}||\mathbf{B}| \cos\theta
\end{aligned}$$

(next page)

Components and Projection

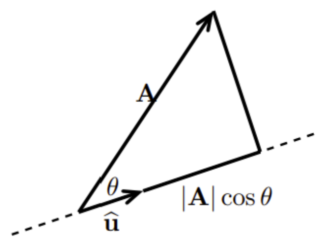
If $\mathbf{A}$ is any vector and $\hat{\mathbf{u}}$ is a unit vector then the *component* of $\mathbf{A}$ in the direction of $\hat{\mathbf{u}}$ is

$$\mathbf{A} \cdot \hat{\mathbf{u}}$$

(Note that the component is a scalar) This can be seen from:

$$\mathbf{A} \cdot \hat{\mathbf{u}} = |\mathbf{A}||\hat{\mathbf{u}}| \cos\theta = |\mathbf{A}| \cos\theta \quad \text{(since } \hat{\mathbf{u}} \text{ is a unit vector)}$$

Geometrically:



The dot product here is the *orthogonal projection* of $\mathbf{A}$ on $\hat{\mathbf{u}}$. The concept of components also extends to non-unit vectors, where the component of $\mathbf{A}$ in the direction of $\mathbf{B}$ is

$$\mathbf{A} \cdot \left(\frac{\mathbf{B}}{|\mathbf{B}|}\right)$$

(Essentially the dot product with the unit vector in the same direction as $\mathbf{B}$)

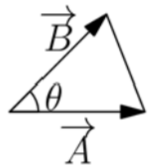
Testing for orthogonality

Since $\cos(\pi/2) = 0$, the dot product gives a test for orthogonality between vectors:

$$\mathbf{A} \perp \mathbf{B} \iff \mathbf{A} \cdot \mathbf{B} = 0$$

B.1.4 Determinant—Intuition

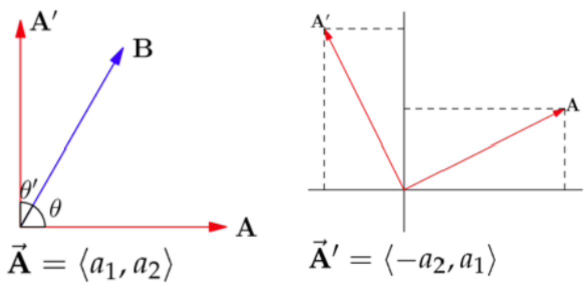
Consider attempting to find the area of a triangle, where $\mathbf{A} = \langle a_1, a_2 \rangle$ and $\mathbf{B} = \langle b_1, b_2 \rangle$:



Intuitively,

$$\begin{aligned}\text{Area} &= \frac{1}{2} \cdot \text{Base} \cdot \text{Height} \\ &= \frac{1}{2}|\mathbf{A}||\mathbf{B}|\sin\theta\end{aligned}$$

now consider an approach where we express this area as a dot product:



Rotating $\mathbf{A}$ by $\pi/2$, notice that

$$\sin\theta = \cos\theta' \quad (\text{where } \theta' = \frac{\pi}{2} - \theta)$$

Since $|\mathbf{A}'| = |\mathbf{A}|$, we get

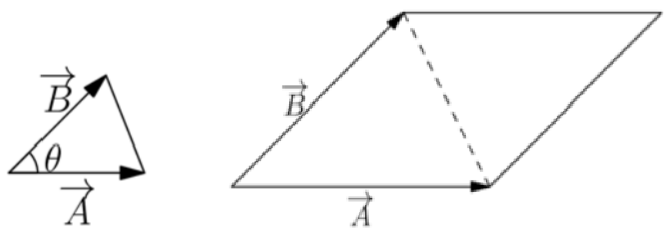
$$\begin{aligned}\frac{1}{2}|\mathbf{A}||\mathbf{B}|\sin\theta &= \frac{1}{2}|\mathbf{A}'||\mathbf{B}|\cos\theta' \\ &= \frac{1}{2}(\mathbf{A}' \cdot \mathbf{B})\end{aligned}$$

(next page)

Notice our expression ($\mathbf{A}' \cdot \mathbf{B}$) simplifies to

$$\begin{aligned}
\mathbf{A}' \cdot \mathbf{B} &= \langle -a_2, a_1 \rangle \cdot \langle b_1, b_2 \rangle \\
&= a_1 b_2 - a_2 b_1
\end{aligned}$$

This absolute value of this result is double the area of our triangle, or the area of a parallelogram formed by **A** and **B**:



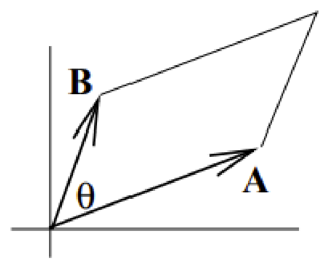
This result is known as the *determinant*:

$$\begin{aligned}
\pm\text{area}(\diamond) &= |\mathbf{A}||\mathbf{B}|\sin\theta \\
&= a_1 b_2 - a_2 b_1 \\
&= \det(\mathbf{A}, \mathbf{B}) \\
\pm\text{area}(\triangleleft) &= \frac{1}{2}|\mathbf{A}||\mathbf{B}|\sin\theta \\
&= \frac{1}{2}\det(\mathbf{A}, \mathbf{B})
\end{aligned}$$

(Note that since $\sin\theta$ can be negative we take the absolute value of the determinant to be the area)

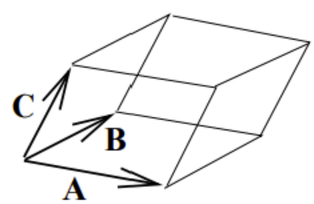
B.1.5 Determinant—Definition

2D determinant



$$\pm\det\begin{pmatrix}a_1 & a_2\\ b_1 & b_2\end{pmatrix} = \pm\begin{vmatrix}a_1 & a_2\\ b_1 & b_2\end{vmatrix}$$
$$= \text{Area of parallelogram with edges } \mathbf{A} = (a_1, a_2),\ \mathbf{B} = (b_1, b_2)$$

3D determinant



$$\pm\det\begin{pmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{pmatrix} = \pm\begin{vmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{vmatrix}$$
$$= \text{Area of parallelepiped with edges row vectors } \mathbf{A}, \mathbf{B}, \mathbf{C}.$$

Generalising, $n \times n$ determinants can be interpreted as the hypervolume in n-space of a n-dimensional parallelotope.

B.1.6 Computing Determinants

In the 2×2 case:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Note that $\left|\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right|$ gives the same answer.

3×3 **determinants**
We first define notation for matrix entries:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \qquad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

- The **ij-entry**, written a_{ij}, is the number in the i-th row and j-th column.
- The **ij-minor**, written $|A_{ij}|$, is the determinant that is left after deleting from $|A|$ the row and column containing a_{ij}.
- The **ij-cofactor**, written A_{ij}, is given as a formula by $A_{ij} = (-1)^{i+j}|A_{ij}|$. Notice that in the case of the 3×3 determinant the sign of the cofactor changes in a checkerboard pattern:

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix}$$

Laplace expansion by cofactors to evaluate a determinant in the 3×3 case occurs as follows: *Select any row (or column) of the determinant. Multiply each entry a_{ij} in that row (or column) by its cofactor A_{ij} and add the three resulting numbers to get the determinant.*

Therefore, Laplace expansion of a third order (3×3) determinant using cofactors of the first row:

$$a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} = |A|$$

using cofactors in the j-th column:

$$a_{1j}A_{1j} + a_{2j}A_{2j} + a_{3j}A_{3j} = |A|$$

Example

$$\begin{aligned} \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a \cdot \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \cdot \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \cdot \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= a(ei - fh) - b(di - fg) + c(dh - eg) \end{aligned}$$

(next page)

Larger determinants
$n \times n$ determinants are evaluated in the same way; the Laplace expansion in the i-th row would be

$$|A| = a_{i1}A_{i1} + a_{i2}A_{i2} + \ldots + a_{in}A_{in}$$

Note that this is an inductive calculation—it expresses the determinant of order n in terms of determinants of order $n-1$.

Example (Laplace expansion by first row)

$$\begin{aligned}
&\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{vmatrix} \\
=&a \cdot \begin{vmatrix} f & g & h \\ j & k & l \\ n & o & p \end{vmatrix} - b \cdot \begin{vmatrix} e & g & h \\ i & k & l \\ m & o & p \end{vmatrix} + c \cdot \begin{vmatrix} e & f & h \\ i & j & l \\ m & n & p \end{vmatrix} - d \cdot \begin{vmatrix} e & f & g \\ i & j & k \\ m & n & o \end{vmatrix} \\
=&\ldots
\end{aligned}$$

The Laplace expansion can be used as the basis of an inductive definition of the $n \times n$ determinant.

Note a few properties of $|A|$ (A is a square array)

- $|A|$ is multiplied by -1 if we interchange two rows or two columns
- $|A| = 0$ if one row or column is all zero, or if two rows or two columns are the same
- $|A|$ is multiplied by c if every element of some row or column is multiplied by c.
- The value of $|A|$ is unchanged if we add to one row (or column) a constant multiple of another row

B.1.7 Cross Product/Vector Product—Definition and formula

Here we describe algebraic and geometric perspectives of the cross product.
Determinant definition for cross product
Given vectors $\mathbf{A} = \langle a_1, a_2, a_3 \rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3 \rangle$, we define the cross product by

$$\begin{aligned}
\mathbf{A} \times \mathbf{B} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \\
&= \mathbf{i} \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - \mathbf{j} \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + \mathbf{k} \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \\
&= (a_2 b_3 - a_3 b_2)\mathbf{i} + (a_3 b_1 - a_1 b_3)\mathbf{j} + (a_1 b_2 - a_2 b_1)\mathbf{k} \\
&= \langle a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1 \rangle
\end{aligned}$$

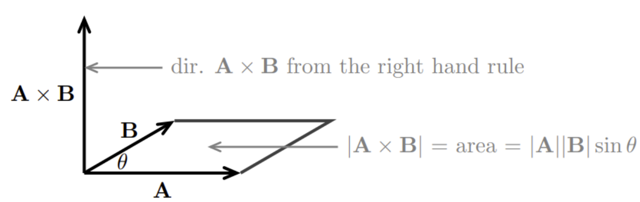
$\mathbf{i}, \mathbf{j}, \mathbf{k}$ are the unit vectors for the x, y, z axes (Note that the top line is technically flawed since we aren't allowed to use vectors as entries in a determinant).
Example

$$\mathbf{i} \times \mathbf{j} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \mathbf{i}(0) - \mathbf{j}(0) + \mathbf{k}(1) = \mathbf{k}$$

Geometric description
The magnitude of $\mathbf{A} \times \mathbf{B}$ is

$$\begin{aligned}
|\mathbf{A} \times \mathbf{B}| &= |\mathbf{A}||\mathbf{B}| \sin\theta \quad (\theta \text{ being the angle between } \mathbf{A} \text{ and } \mathbf{B}) \\
&= \text{Area of parallelogram spanned by } \mathbf{A} \text{ and } \mathbf{B}
\end{aligned}$$

The direction of $\mathbf{A} \times \mathbf{B}$ is perpendicular to the plane of $\mathbf{A}$ and $\mathbf{B}$. The directon is given by the *right hand rule*—pointing one's fingers in the direction of $\mathbf{A}$ and curling them toward $\mathbf{B}$, one's thumb would point in the direction of $\mathbf{A} \times \mathbf{B}$:



(next page)

Note therefore, a few properties of the cross product:

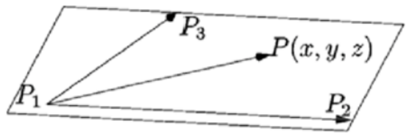
- $\mathbf{A} \times \mathbf{A} = \mathbf{0}$
- $\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}$ (Anti-commutivity)
- $\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}$ (Distributive)
- $(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} \neq \mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ (Non-associativity)

For the unit vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$, we have

$$\mathbf{i} \times \mathbf{j} = \mathbf{k}, \quad \mathbf{j} \times \mathbf{k} = \mathbf{i}, \quad \mathbf{k} \times \mathbf{i} = \mathbf{j}$$

B.1.8 Equations of Planes I

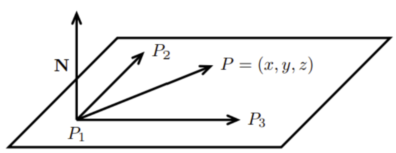
Intuition
Consider a plane made up of three points P_1, P_2, P_3. We want to determine whether a point $P = (x, y, z)$ is in the plane:



One approach would be to compute the determinant of the three vectors formed by the four points; since if P were within the plane the volume of the parallelepiped formed by the three points would be zero. Thus P is within the plane if

$$\det(\overrightarrow{\mathbf{P_1P_2}}, \overrightarrow{\mathbf{P_1P_3}}, \overrightarrow{\mathbf{P_1P}}) = 0$$

Another approach would be to find the *normal vector* $\overrightarrow{\mathbf{N}}$ to the plane; where if $\overrightarrow{\mathbf{P_1P}}$ is orthogonal to $\overrightarrow{\mathbf{N}}$, then P lies within the plane:



We find $\overrightarrow{\mathbf{N}}$ using the cross product:

$$\overrightarrow{\mathbf{N}} = \overrightarrow{\mathbf{P_1P_2}} \times \overrightarrow{\mathbf{P_1P_3}}$$

Orthogonality can then be tested using the dot product, giving us

$$\overrightarrow{\mathbf{P_1P}} \cdot \overrightarrow{\mathbf{N}} = \overrightarrow{\mathbf{P_1P}} \cdot (\overrightarrow{\mathbf{P_1P_2}} \times \overrightarrow{\mathbf{P_1P_3}}) = 0$$

if P is within the plane. It turns out that this expression, called the *triple product*, is equal to the determinant:

$$\begin{aligned}
b \times c &= (b_2c_3 - b_3c_2, b_3c_1 - b_1c_3, b_1c_2 - b_2c_1) \\
\underbrace{a \cdot (b \times c)}_{\text{triple product}} &= a_1b_2c_3 - a_1b_3c_2 + a_2b_3c_1 - a_2b_1c_3 + a_3b_1c_2 - a_3b_2c_1 \\
&= \det\left(\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}\right)
\end{aligned}$$

(next page)

Example
We can use this to find the equation of a plane. Consider a plane containing three points $P_1 = (1,3,1), P_2 = (1,2,2), P_3 = (2,3,3)$. We have two vectors $\overrightarrow{\mathbf{P_1P_2}}$ and $\overrightarrow{\mathbf{P_1P_3}}$ in the plane, so we have

$$\overrightarrow{\mathbf{N}} = \overrightarrow{\mathbf{P_1P_2}} \times \overrightarrow{\mathbf{P_1P_3}} = (-2,1,1)$$

Since for any point P in the plane $\overrightarrow{\mathbf{P_1P}}$ is orthogonal to $\mathbf{N}$, we have

$$\begin{aligned}
\overrightarrow{\mathbf{N}} \cdot \overrightarrow{\mathbf{P_1P}} &= 0 \\
(-2,1,1) \cdot (x-1, y-3, z-1) &= 0 \\
-2x + y + z &= 2
\end{aligned}$$

B.1.9 Matrix Operations

1. Scalar multiplication
Multiplying each entry by a scalar

$$cA = (ca_{ij})$$

2. Matrix addition
Addition of corresponding entries (as such both matrices have to be of the same size)

$$A + B = (a_{ij} + b_{ij})$$

3. Transposition
The *transpose* of the $m \times n$ matrix A is the $n \times m$ matrix obtained by making the rows of A the columns of the new matrix, denoted by A^T:

$$\text{If} \quad A = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}, \quad A^T = \begin{pmatrix} a & d \\ b & e \\ c & f \end{pmatrix}$$

$$\text{If} \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad A^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

4. Matrix multiplication
Schematically, we have

$$\underbrace{A}_{m\times n} \cdot \underbrace{B}_{n\times p} = \underbrace{C}_{m\times p}$$

$$\text{where} \quad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

For the multiplication to be defined, A must have as many *columns* as B has *rows*. The ij-th entry of the product matrix C is the dot product of the i-th row of A with the j-th column of B.
(next page)

Properties of matrix multiplication

- Distributive: $A(B+C) = AB + AC, \quad (A+B)C = AC + BC$
- Associative: $(AB)C = A(BC)$
- Given *identity* matrix I, $AI = IA$
- In general, $AB \neq BA$ (non-commutative). (There are some important exceptions, such as in the case with the identity matrix)
- For two square $n \times n$ matrices A and B. We have the *determinant law*:

$$|AB| = |A||B| \quad \text{or} \quad \det(AB) = \det(A)\det(B)$$

Extraction of single row or column A useful fact is that a simple row or column vector can be used to pick out a row or column of a given matrix as follows:

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} b \\ e \\ h \end{pmatrix} \quad \text{(Middle column)}$$

$$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \begin{pmatrix} a & b & c \end{pmatrix} \quad \text{(First row)}$$

B.1.10 Meaning of Matrix Multiplication

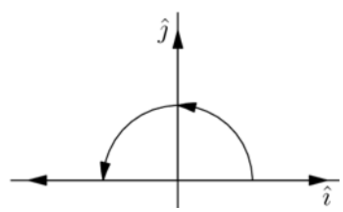
Matrices as transformations
Given two matrices A and B, AB can be seen as representing:

Do transformation B, then transformation A

Matrix multiplication is done *right to left*; with the first multiplication representing a transformation from the basic unit vectors.
Example
The matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ represents a plane rotation by 90° counter clockwise; consider how multiplication changes the unit vectors:

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Visualised:



$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix}$$

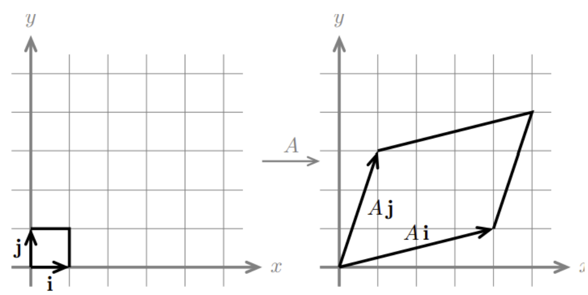
(next page)

More intuition about determinants

The matrix $A = \left(\begin{smallmatrix} 4 & 1 \\ 1 & 3 \end{smallmatrix}\right)$ transforms the unit square into a parallelogram as follows:

$$\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \qquad \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Notice the transformation of the unit vectors leads to new vectors with endpoints specified in A:



The resulting parallelogram has edges that are identical to the entries of A.

Notice that the area of this parallelogram can be found therefore using

$$\text{Area of resulting parallelogram} = \det(A)$$

illustrating a relationship between a change in area after a transformation and the corresponsting determinant.

B.1.11 Matrix Inverses—Intuition

Consider a linear system of equations:

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3
\end{aligned}$$

In such systems a_{ij} and b_i are given, and we want to solve for x_i. We can use matrix multiplication to abbreviate the system by

$$A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Where A is the square matrix of coefficients (a_{ij}). Notice that this works for any $n \times n$ system of equations. Given $\mathbf{b}$ and A, we want to solve for $\mathbf{x}$.

With that, suppose we could find a square matrix M, the same size as A, such that

$$MA = I \quad (I \text{ being the identity matrix})$$

since that would allow for

$$\begin{aligned}
A\mathbf{x} &= \mathbf{b} \\
M(A\mathbf{x}) &= M\mathbf{b} \\
\mathbf{x} &= M\mathbf{b}
\end{aligned}$$

This works because

$$\begin{aligned}
M(A\mathbf{x}) &= (MA)\mathbf{x} \quad \text{(associative law)} \\
&= I\mathbf{x} \\
&= \mathbf{x}
\end{aligned}$$

B.1.12 Inverse Matrix Computation

Here we outline a method to computing inverse matrices by hand

On the existence of an inverse
Note that the inverse M of matrix A doesn't always exist:

$$M \text{ exists} \iff |A| \neq 0$$

This follows from the determinant law for square matrices:

$$\det(AB) = \det(A)\det(B)$$

since

$$MA = I \implies |MA| = |I| = 1 \implies |A| \neq 0$$

(the implication in the other direction requires more)

Definition
Let A by an $n \times n$ matrix, with $|A| \neq 0$. Then the **inverse** of A is an $n \times n$ matrix A^{-1}, such that

$$A^{-1}A = I_n, \qquad AA^{-1} = I_n$$

(it is enough to verify either equation; the other follows automatically.) This gives us

$$|A| \neq 0 \implies \text{the unique solution of } A\mathbf{x} = \mathbf{b} \text{ is } \mathbf{x} = A^{-1}\mathbf{b}$$

(next page)

Calculating the inverse of a 3×3 matrix
We consider the 3×3 case; let A be the matrix

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

The formulas for its *inverse* A^{-1} and for an auxiliary matrix $\operatorname{adj} A$ called the *adjoint/adjugate* of A are

$$A^{-1} = \frac{1}{|A|} \operatorname{adj} A = \frac{1}{|A|} \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}^T$$

Recall that A_{ij} is the *cofactor* of the element a_{ij} in the matrix (which is its *minor* with its sign changed by the checkerboard rule). A procedure to calculate inverse matrices is therefore:

1. Calculate the matrix of minors.
2. Change the signs of the matrix according to the corresponding cofactors (checkerboard rule).
3. Transpose the resulting matrix (this gives $\operatorname{adj} A$).
4. Divide every entry by $|A|$.

Notice that this formula, as expected, doesn't work for $|A| \neq 0$, and that each A_{ij} is a signed determinant (which isn't immediately obvious from the notation).

Illustrated:

$$\underset{\text{matrix}}{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \to \underset{\text{cofactor matrix}}{\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}} \underset{\mathrm{T}}{\to} \underset{\operatorname{adj} A}{\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}} \underset{\frac{1}{|A|}}{\to} \underset{\text{inverse of } A}{\frac{1}{|A|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}}$$

Example
Given matrix A:

$$\underset{\text{matrix } A}{\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}} \to \underset{\text{cofactor matrix}}{\begin{pmatrix} 1 & 1 & -1 \\ 0 & 2 & 0 \\ 1 & -1 & 1 \end{pmatrix}} \underset{\mathrm{T}}{\to} \underset{\operatorname{adj} A}{\begin{pmatrix} 1 & 0 & 1 \\ 1 & 2 & -1 \\ -1 & 0 & 1 \end{pmatrix}} \underset{\frac{1}{|A|}}{\to} \underset{\text{inverse of } A}{\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}}$$

(next page)

Intuition
We want to show that $A \cdot A^{-1} = I$ or equivalently, $A{\cdot}\operatorname{adj} A = |A|I$. This looks like

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix} = \begin{pmatrix} |A| & 0 & 0 \\ 0 & |A| & 0 \\ 0 & 0 & |A| \end{pmatrix}$$

Consider the first entry of the first row in the resulting matrix. We want to show that

$$a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} = |A|$$

Notice however that this already matches the formula for the determinant of $|A|$:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

Now consider the second entry of the first row; we want to show that

$$a_{11}A_{21} + a_{12}A_{22} + a_{13}A_{23} = 0$$

This sum would represent the determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

Notice that the determinant here will equal 0 since two of its rows are the same. One can also visualise a parallelepiped formed by the column vectors and see it has no volume.

B.1.13 Equations of Planes II

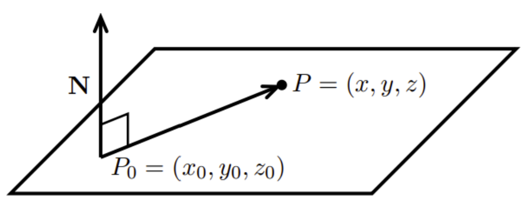
Planes in point-normal form

The basic data determining a plane is a point P_0 in the plane and a vector N orthogonal to the plane. (Recall N is said to be the *normal* to the plane)

Given $P_0 = (x_0, y_0, z_0)$ and $\overrightarrow{\mathbf{N}} = \langle a, b, c\rangle$, we want the equation of a plane; letting $P = (x, y, z)$ be an arbitrary point in the plane, then the vector $\overrightarrow{\mathbf{P_0P}}$ is in the plane—making it orthogonal to $\mathbf{N}$—giving us

$$\begin{aligned} &\mathbf{N} \cdot \overrightarrow{\mathbf{P_0P}} = 0 \\ \iff &\langle a, b, c\rangle \cdot \langle x - x_0, y - y_0, z - z_0\rangle = 0 \\ \iff &a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \end{aligned}$$

this last line is called the *point-normal form* for the plane (notice how the entries of the normal vector can be seen in the equation)

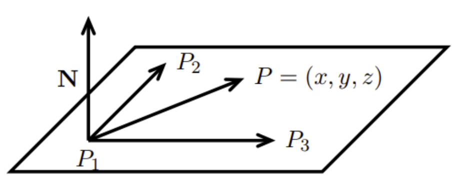


Recall that $\mathbf{N}$ can be found using the cross product of any two vectors in the plane. (Also notice that any scalar multiple of $\mathbf{N}$ is still orthogonal to the plane— one plane can be represented by different equations.)

Example: Find the plane containing $P_1 = (1, 2, 3), P_2 = (0, 0, 3), P_3 = (2, 5, 5)$

$$\mathbf{N} = \overrightarrow{\mathbf{P_1P_2}} \times \overrightarrow{\mathbf{P_1P_3}} = \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -1 & -2 & 0 \\ 1 & 3 & 2 \end{pmatrix} = \langle -4, 2, 1\rangle$$

Expressed in point normal form:



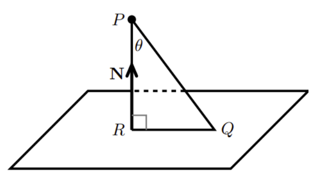
$$-4(x - 1) + 2(y - 2) - (z - 3) = 0$$

B.1.14 Distance between lines and planes

Point to plane
Given a point P and a plane with normal $\overrightarrow{\mathbf{N}}$ containing a point Q, the distance d from P to the plane is

$$d = |\overrightarrow{\mathbf{PQ}}| \cos\theta = \left| \overrightarrow{\mathbf{PQ}} \cdot \frac{\overrightarrow{\mathbf{N}}}{|\mathbf{N}|} \right|$$

this is seen intuitively from the geometric definition of the dot product:

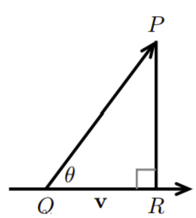


the length is equivalent to the orthogonal projection of $\overrightarrow{\mathbf{PQ}}$ onto the unit vector in the direction of $\overrightarrow{\mathbf{N}}$. $\square$

Point to line (alternatively)
Given a point P and a line with direction vector $\mathbf{v}$ containing point Q, the distance d from P to the line is given by

$$d = \overrightarrow{\mathbf{QP}} \sin\theta = \left| \overrightarrow{\mathbf{QP}} \times \frac{\mathbf{v}}{|\mathbf{v}|} \right|$$

Notice d can be found using $\overrightarrow{\mathbf{QP}} \sin\theta$:



using the relation $|\mathbf{A} \times \mathbf{B}| = |\mathbf{A}||\mathbf{B}| \sin\theta$

$$\overrightarrow{\mathbf{QP}} \sin\theta = \left| \overrightarrow{\mathbf{QP}} \times \frac{\mathbf{v}}{|\mathbf{v}|} \right| \quad \square$$

(next page)

Distance between parallel planes
The distance between two planes can simply be seen as the distance between from a point to a plane:

Example: Find the distance between the planes $x+2y-z=4$ and $x+2y-z=3$

Notice from the equations that both planes have the normal $\mathbf{N} = \langle 1, 2, -1\rangle$. The orthogonal projection of any vector from one plane to the other onto a unit vector in the same direction as $\mathbf{N}$ is equivalent to the length between the planes. Choosing $P = (4, 0, 0)$ on the first plane and $Q = (1, 0, 0)$ on the second plane,

$$d = \left| \overrightarrow{\mathbf{QP}} \cdot \frac{\overrightarrow{\mathbf{N}}}{|\mathbf{N}|} \right| = \frac{1}{\sqrt{6}} |\langle 3, 0, 0\rangle \cdot \langle 1, 2, -1\rangle| = \frac{\sqrt{6}}{2} \quad \square$$

Distance between skew lines
We place the lines in parallel planes and find the distance between the planes as in the previous example. As usual it's easy to find a point on each line; $\mathbf{N}$ can be found using

$$\mathbf{N} = \mathbf{v_1} \times \mathbf{v_2}$$

where $\mathbf{v_1}$ and $\mathbf{v_2}$ are the direction vectors of the lines.

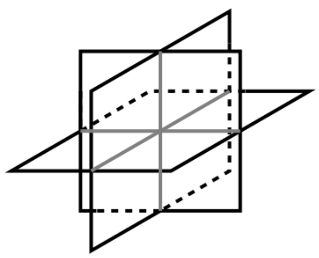
B.1.15 Geometry of 3×3 systems

Consider, geometrically, the possible solutions to a 3×3 system such as

$$\begin{array}{rcrcrcr} 6x & + & 5y & + & 3z & = & 1 \\ x & + & 2y & + & 7z & = & 4 \\ 2x & - & 2y & - & 2z & = & 8 \end{array}$$

Each equation is one of a plane. Geometrically solving the system means finding the intersection of three planes (the set of points that lie on all three planes).

Usually three planes intersect at a point:

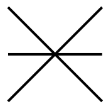


However, other possibilities exist. Altogether there are four:

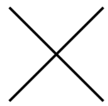
1. Intersection at a point (1 solution)
2. Intersection in a line (∞ solutions)
   (a) 3 different planes, third plane contains the intersection of the first two.
   (b) 2 planes are the same, the third plane intersects them in a line.
3. Intersection in a plane (∞ solutions)
   (a) All three planes are the same
4. No intersections (0 solutions)
   (a) The planes are different, but are parallel
   (b) Two planes are parallel, the third crosses them
   (c) The planes are different and none are parallel, but the lines of intersection of each pair are parallel.
   (d) Two planes are same and parallel to the third

(next page)

Each different case illustrated (imagining the planes as extending vertically out of the page)

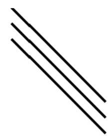


Case (2a)

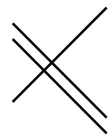


Case (2b)

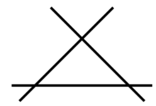
Case (3a)



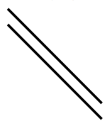
Case (4a)



Case (4b)



Case (4c)



Case (4d)

B.1.16 Solutions to Square Systems

Theorems for homogeneous and inhomogeneous systems
The linear system $A\mathbf{x} = \mathbf{b}$ is *homogeneous* if $\mathbf{b} = \mathbf{0}$; otherwise it is *inhomogeneous.*

Theorem 1. *Let A be a $n \times n$ matrix.*

$$|A| \neq 0 \implies A\mathbf{x} = \mathbf{b} \quad \textit{has the unique solution,} \quad \mathbf{x} = A^{-1}\mathbf{b}$$
$$|A| \neq 0 \implies A\mathbf{x} = \mathbf{0} \quad \textit{has only the trivial solution,} \quad \mathbf{x} = \mathbf{0}$$

Note that the trivial solution in this case is the *unique* solution for the case where $\mathbf{b} = \mathbf{0}$. Often it is stated and used in the contrapositive form:

$$A\mathbf{x} = \mathbf{0} \quad \text{has a non-zero solution} \implies |A| = 0$$

(The contrapositive of a statement $P \implies Q$ is not-$Q \implies$ not-P; the two statements say the same thing)

Theorem 2. *Let A be a $n \times n$ matrix*

$$|A| = 0 \implies A\mathbf{x} = \mathbf{0} \quad \textit{has non-trivial (non-zero) solutions}$$
$$|A| = 0 \implies A\mathbf{x} = \mathbf{b} \quad \textit{usually has no solutions, but has solutions for some } \mathbf{b}$$

In this last case, we call the system *consistent* if it has solutions, and *inconsistent* otherwise.

In summary, for **Homogeneous systems**:

$$A\mathbf{x} = \mathbf{0} \quad \text{has a non-trivial solution} \iff |A| = 0$$

For **Inhomogeneous systems**:

$$\text{If} \quad |A| \neq 0, \quad A\mathbf{x} = \mathbf{b} \quad \text{has a unique solution for} \quad \mathbf{x} = A^{-1}\mathbf{b}$$
$$\text{If} \quad |A| = 0, \quad A\mathbf{x} = \mathbf{b} \quad \text{usually has no solutions, except for some } \mathbf{b}$$

(next page)

Intuition for theorem 2
First we look at

$$|A| = 0 \implies A\mathbf{x} = \mathbf{0} \quad \textit{has non-trivial (non-zero) solutions}$$

We represent the three rows of A by the row vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and we let $\mathbf{x} = (x, y, z)$. We have

$$A\mathbf{x} = \mathbf{0} \quad \text{is the same as the system} \quad \mathbf{a} \cdot \mathbf{x} = 0, \quad \mathbf{b} \cdot \mathbf{x} = 0, \quad \mathbf{c} \cdot \mathbf{x} = 0$$

Recall that

$$|A| = \mathbf{a} \cdot \mathbf{b} \times \mathbf{c} = \text{volume of parallelepiped spanned by } \mathbf{a}, \mathbf{b}, \mathbf{c}.$$

if $|A| = 0$, the parallelepiped formed (by $\mathbf{a}, \mathbf{b}, \mathbf{c}$) has 0 volume. This means that the origin vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ lie in a plane; any non-zero vector x orthogonal to the plane will also be orthogonal to $\mathbf{a}, \mathbf{b}, \mathbf{c}$, and will therefore be a solution to the system, leading to non-trivial solutions.

Now we look at

$$\text{If} \quad |A| = 0, \quad A\mathbf{x} = \mathbf{b} \quad \text{usually has no solutions, except for some } \mathbf{b}$$

Similar to earlier, $A\mathbf{x} = \mathbf{d}$, $\mathbf{d}$ being a column vector $\mathbf{d} = (d_1, d_2, d_3)^T$, is the same as the system

$$\mathbf{a} \cdot \mathbf{x} = d_1, \quad \mathbf{b} \cdot \mathbf{x} = d_2 \quad \mathbf{c} \cdot \mathbf{x} = d_3$$

If $|A| = 0$, the three origin vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ lie in a plane. This means we can write one of them, say $\mathbf{c}$, as a *specific* linear combination of $\mathbf{a}$ and $\mathbf{b}$:

$$\mathbf{c} = r\mathbf{a} + s\mathbf{b}, \quad \text{where } r, s \text{ are real numbers}$$

x being some vector, it follows that

$$\begin{aligned} \mathbf{c} \cdot \mathbf{x} &= r\mathbf{a} \cdot \mathbf{x} + s\mathbf{b} \cdot \mathbf{x} \\ d_3 &= rd_1 + sd_2 \end{aligned}$$

unless the components of $\mathbf{d}$ satisfy this relationsip (where one entry is a specific linear combination of the other two), there cannot be a solution; thus in general there are non solutions.

In the case where $\mathbf{d}$ does satisfy this relation, then we get a system of two equations with three unknowns, which in general will have a non-zero solution unless they represent two parallel planes.

Singular and Nonsingular matrices

We say that a square matrix A is *singular* if $|A| = 0$ and *nonsingular* or *invertible* if $|A| \neq 0$. This comes from the fact that A^{-1} exists if and only if $|A| \neq 0$.

Computational difficulties

Even if A is nonsingular, the solution of $A\mathbf{x} = \mathbf{b}$ becomes difficult to compute should $|A| \approx 0$, where A is said to be *almost-singular*. In this case (since $|A|$ occurs in the denominator in the formula for A^{-1}) the solutions tend to be sensitive to small changes in the entries of A (the coefficients of the equations). Systems behaving like this are said to be *ill-conditioned*; they are difficult to solve numerically and require special methods.

Intuitively, consider a 2×2 system, which represents a pair of lines. If $|A| \approx 0$ and its entries are not small, then its two rows must be vectors which are almost parallel (they span a parallelogram of small area). Although their intersection point exists, it is highly sensitive to the exact positions of the two lines (the values of the coefficients of the system of equations).

B.2 Parametric Equations

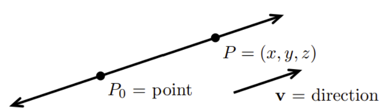
B.2.1 Parametric Equations of lines

Parametric curves are the idea of a point moving in space tracing out a path over time. In three spatial dimensions this gives us four variables to consider—position (x, y, z) and the independent variable t (which we can think of as time). They are written as

$$x = x(t), \quad y = y(t), \quad z = z(t)$$

to indicate that x, y, z are functions of t. t is the parameter, and x, y, z are the *parametric equations*.

Suppose our point moves on a line. The basic data we need to specify a line are a point on the line and a vector parallel to the line (intuitively, this represents a point and a direction):



Example: Given a line through $P_0 = (1, 2, 3)$, parallel to $\mathbf{v} = \langle 1, 3, 5\rangle$:
If $P = (x, y, z)$ is on the line then

$$\overrightarrow{\mathbf{P_0P}} = \langle x-1, y-2, z-3\rangle$$

is parallel to $\langle 1, 3, 5\rangle$; that is, $\overrightarrow{\mathbf{P_0P}}$ is a scalar multiple of $\langle 1, 3, 5\rangle$. We call this scale t and write

$$\begin{aligned} &\langle x-1, y-2, z-3\rangle = t\langle 1, 3, 5\rangle \\ \iff\ & x-1 = t, \quad y-2 = 3t, \quad z-3 = 5t \\ \iff\ & x = 1+t, \quad y = 2+3t, \quad z = 3+5t \end{aligned}$$

Generally, a line through $P_0 = (x_0, y_0, z_0)$ in the direction of $\mathbf{v} = \langle v_1, v_2, v_3\rangle$ has parameterization

$$\begin{aligned} &\langle x, y, z\rangle = \langle x_0 + tv_1, y_0 + tv_2, z_0 + tv_3\rangle \\ \iff\ & x = x_0 + tv_1, \quad y = y_0 + tv_2, \quad z = z_0 + tv_3 \end{aligned}$$

intuivively, the starting point plus its movement based on t. (next page)

Application: Intersection of a line and a plane
Example: Given the plane $\mathcal{P} = 2x+y-4z = 4$, find all the points of intersection with the line $x = t, y = 2 + 3t, z = t$.

By substituting the formulas for x, y, z into the equation and solving for t:

$$2(t) + (2 + 3t) - 4(t) = 4 \iff t = 2$$

using $t = 2$ we get our point $(x, y, z) = (2, 8, 2)$.

Consider another point $x = 1 + t, y = 4 + 2t, z = t$; substituting gives

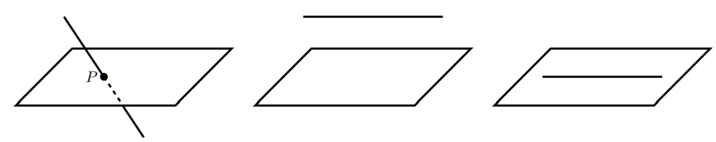
$$2(1+t)+(4+2t)-4(t) = 4 \iff 6 = 4 \iff \text{no values of } t \text{ satisfy this equation}$$

There are no points of intersection.

Finally consider $x = t, y = 4 + 2t, z = t$; substituting:

$$2(t) + (4 + 2t) - 4(t) = 4 \iff 4 = 4 \iff \text{all values of } t \text{ satisfy this equation}$$

the line is contained in the plane, thus all points of the line intersect with the plane. Each case illustrated:



(a) line intersects the plane in a point　(b) line is parallel to the plane　(c) line is in the plane

B.2.2 Parametric Curves

Now we consider parametric equations of more general trajectories. We define the *position vector*, which is just the vector from the origin to the moving point, as

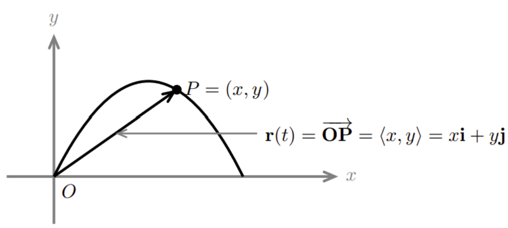
$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} = \langle x(t), y(t), z(t)\rangle$$

this is just the vector from the origin to the moving point.

Example
A rocket fired from the origin with initial x-velocity $v_{0,x}$ and initial y-velocity $v_{0,y}$ has a trajectory corresponding to the parametric equation

$$x(t) = v_{0,x}t, \quad y(t) = -\frac{1}{2}gt^2 + v_{0,y}t$$

At time t the rocket is at point $P = (x(t), y(t))$, the position vector is $r(t) = \overrightarrow{\mathbf{OP}}$:



Circles
The parametric curve in the plane

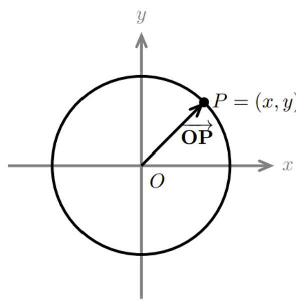
$$x(t) = a\cos t, \quad y(t) = a\sin t$$

can be further expressed as

$$x^2 + y^2 = a^2\cos^2 t + a^2\sin^2 t = a^2$$

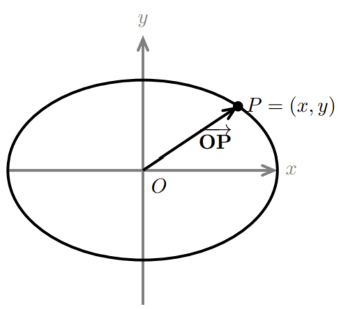
Therefore the trajectory describes a circle of radius a centered at O.
(next page)

We call $x(t) = a\cos t, y(t) = a\sin t$ the *parametric form* of the curve and $x^2 + y^2 = a^2$ the *symmetric form*.



Ellipse:
The circle is easily changed to an ellipse by

$$\begin{aligned} \text{Parametric form:} \quad & x(t) = a\cos t, y(t) = b\cos t \\ \text{Symmetric form:} \quad & \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \end{aligned}$$



B.2.3 Velocity and Acceleration, Product rule for vector derivatives

Considering a position vector

$$\boldsymbol{r}(t) = \langle x(t), y(t), z(t) \rangle$$

The *velocity* vector is given by

$$\boldsymbol{v} = \frac{d\boldsymbol{r}}{dt} = \left\langle \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right\rangle$$

and the *acceleration* vector is given by

$$\boldsymbol{a} = \frac{d\boldsymbol{v}}{dt}$$

Product rule for vector derivatives (dot product)
We show that for two position vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, we have

$$\frac{d(\boldsymbol{r}_1 \cdot \boldsymbol{r}_2)}{dt} = \boldsymbol{r}_1' \cdot \boldsymbol{r}_2 + \boldsymbol{r}_1 \cdot \boldsymbol{r}_2'$$

See that

$$\begin{aligned}
\boldsymbol{r}_1 \cdot \boldsymbol{r}_2 &= x_1 x_2 + y_1 y_2 \\
\frac{d(\boldsymbol{r}_1 \cdot \boldsymbol{r}_2)}{dt} &= \frac{d(x_1 x_2 + y_1 y_2)}{dt} \\
&= x_1' x_2 + x_1 x_2' + y_1' y_2 + y_1 y_2' \\
&= (x_1' x_2 + y_1' y_2) + (x_1 x_2' + y_1 y_2') \\
&= \boldsymbol{r}_1' \cdot \boldsymbol{r}_2 + \boldsymbol{r}_1 \cdot \boldsymbol{r}_2'
\end{aligned}$$

Product rule for vector derivatives (cross product)
By expansion of $(\boldsymbol{r}_1 \times \boldsymbol{r}_2)$, same as in the case of the dot product above, one can prove

$$\frac{d(\boldsymbol{r}_1 \times \boldsymbol{r}_2)}{dt} = \boldsymbol{r}_1' \times \boldsymbol{r}_2 + \boldsymbol{r}_1 \times \boldsymbol{r}_2'$$

B.2.4 Velocity and Arc length

Arc length
The arc length s is the distance travelled along the trajectory of a curve; since *speed* is just the magnitude of velocity and a measure of arc length per unit time, we have

$$\frac{ds}{dt} = \text{speed} = |\boldsymbol{v}|$$

intuitively, the arc length is just a magnitude—it doesn't give information on the direction of the curve. (as one can see from the equation, arc length can be obtained from integrating $|\boldsymbol{v}|$ over time)

Unit tangent vector
The *unit tangent vector* gives a measure of the 'trajectory' of the curve. It is given by

$$\hat{\boldsymbol{T}} = \frac{\boldsymbol{v}}{|\boldsymbol{v}|}$$

essentially a unit vector (of $\boldsymbol{v}$) that is a tangent to the curve at a particular point.

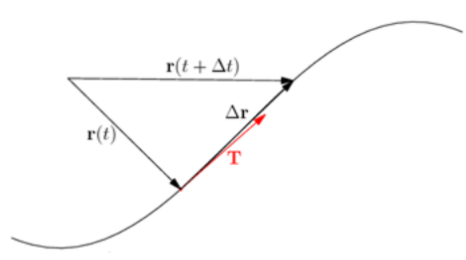
Velocity as a combination of the two
Consider the formula for velocity combined with the chain rule:

$$\boldsymbol{v} = \frac{d\boldsymbol{r}}{ds}\underbrace{\frac{ds}{dt}}_{|\boldsymbol{v}|} \qquad \Longrightarrow \qquad \frac{d\boldsymbol{r}}{ds} = \hat{\boldsymbol{T}}$$

Intuitively, $\boldsymbol{v}$ can be broken up as

$$\text{Velocity has} \begin{cases} \text{direction} : \hat{\boldsymbol{T}} \\ \text{length} : \text{speed} = \frac{ds}{dt} \end{cases}$$

and we have



$$\boldsymbol{v} = \frac{d\boldsymbol{r}}{dt} = \hat{\boldsymbol{T}}\frac{ds}{dt}$$

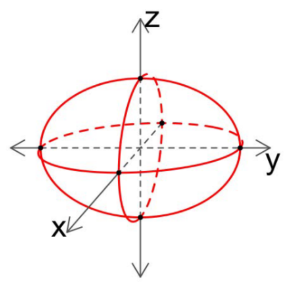
B.3 Partial Derivatives

B.3.1 Graphing functions of two variables

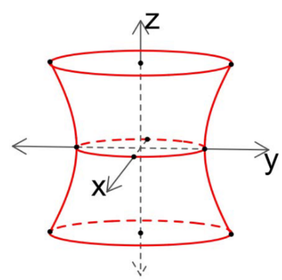
Consider a function that takes two inputs x and y, which can be written as

$$z = f(x, y)$$

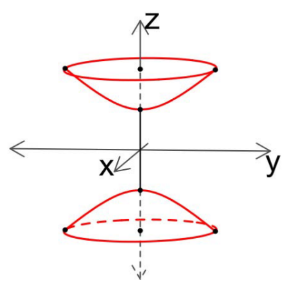
in this case we say that x and y are independent variables and z is a dependent variable. Consider the graphs of a few of such functions:



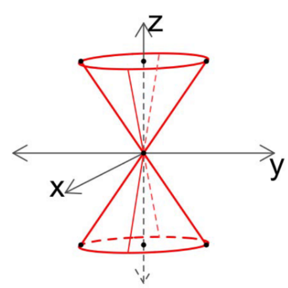
Ellipsoid: $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$



Hyperboloid of one sheet: $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$



Hyperboloid of two sheets: $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1$



Elliptic cone: $\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z^2}{c^2}$

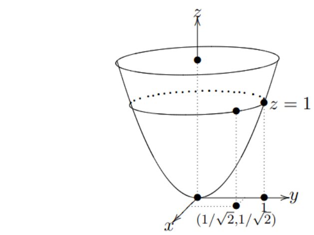
In some cases graphs plotted like these can be hard to visualise.

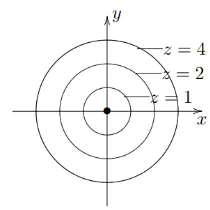
(next page)

Contour plots

Level curves and *contour plots* are another way of visualising functions of two variables; where we plot a 'top down' view of the graph using *contours*, which are curves at fixed heights $z =$constant.

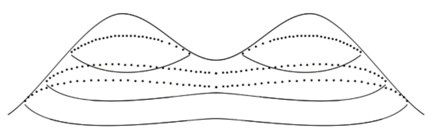
To illustrate this, consider the plot of $z = x^2 + y^2$:



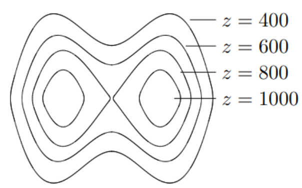


The level curve at height $z = 1$ is the circle $x^2 + y^2 = 1$.

Another example:



Mountain pass



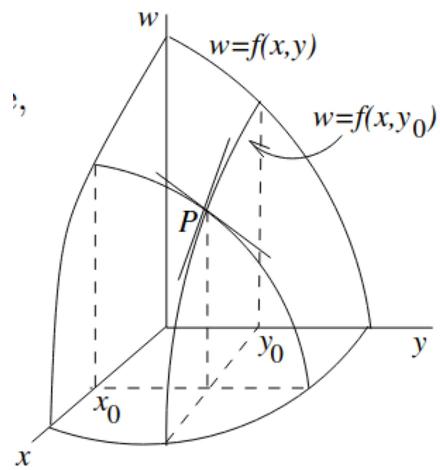
Level curves

B.3.2 Partial Derivatives

Consider a function of two variables $w = f(x, y)$, fixing one variable $y = y_0$ and letting the other x vary, we get the function for *one* variable,

$$w = f(x, y_0), \quad \text{the } \textit{partial function} \text{ for } y = y_0$$

(a similar graph $w = f(x_0, y)$ can be obtained should we fix the other variable instead) See that its graph is confined to the vertical plane $y = y_0$:



The slope of this graph at point P where $x = x_0$ is given by the derivative

$$\frac{d}{dx} f(x, y_0)\bigg|_{x_0}, \quad \text{or} \quad \frac{\partial f}{\partial x}\bigg|_{(x_0, y_0)}, \quad \text{or} \quad \left(\frac{\partial f}{\partial x}\right)_0, \quad \text{or} \quad \left(\frac{\partial w}{\partial x}\right)_0$$

This is called the *Partial Derivative* of f with respect to x at the point (x_0, y_0), written in the standard notation on the right and middle. The partial derivative is just the ordinary derivative of the partial function

Similarly the partial derivative of f with respect to y, where we fix $x = x_0$, is written as

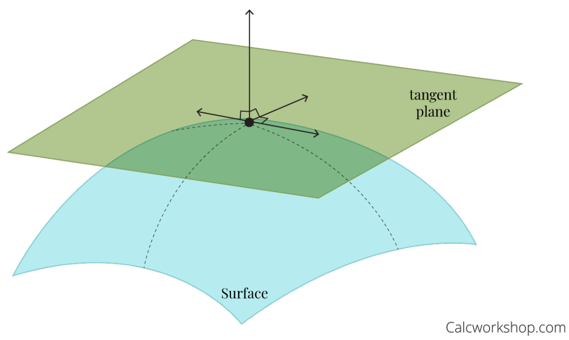
$$\frac{\partial f}{\partial y}\bigg|_{(x_0, y_0)}, \quad \text{or} \quad \left(\frac{\partial f}{\partial y}\right)_0, \quad \text{or} \quad \left(\frac{\partial w}{\partial y}\right)_0$$

More variables
For functions for three or more variables $w = f(x, y, z, \ldots)$, we can't quite visualise them with graphs, but the idea behind partial differentiation remains the same: to define the partial derivative with respect to x, hold all other variables constant and take the ordinary with respect to x.

B.3.3 Approximation Formula

The tangent plane
For functions of single variables $w = f(x)$, the tangent line to its graph at a point (x_0, w_0), is the line passing through (x_0, w_0) with slope $\left(\frac{dw}{dx}\right)_0$.

For a function of two variables $w = f(x, y)$, the natural ananalogue is the *tangent plane* to the graph at a point:



Intuitively, the plane

- must pass through (x_0, y_0, w_0), where $w_0 = f(x_0, y_0)$, and
- must contain the tangent lines to the graphs of the two partial functions at (x_0, y_0, w_0).

First we have the equations for the tangent lines to the two partial functions;

$$\text{If} \quad \frac{\partial f}{\partial x}(x_0, y_0) = a, \quad \text{then} \quad L_1 = \begin{cases} w = w_0 + a(x - x_0) \\ y = y_0 \end{cases}$$

$$\text{If} \quad \frac{\partial f}{\partial y}(x_0, y_0) = b, \quad \text{then} \quad L_2 = \begin{cases} w = w_0 + b(y - y_0) \\ x = x_0 \end{cases}$$

where L_1 and L_2 are both tangent to the graph w at point (x_0, y_0, w_0); together they give the equation for the tangent plane as

$$w = w_0 + \left(\frac{\partial w}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial w}{\partial y}\right)_0 (y - y_0)$$

($\left(\frac{\partial w}{\partial x}\right)_0$ and $\frac{\partial f}{\partial x}(x_0, y_0)$ mean the same thing, just a change in notation)

(next page)

Approximation Formula

The intuitive idea is that if we stay near (x_0, y_0, w_0) , the graph of the tangent plane would be a good approxmation of the graph of the function $w = f(x, y)$:

$$\underbrace{w = f(x,y)}_{\text{height of graph}} \approx \underbrace{w_0 + \left(\frac{\partial w}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial w}{\partial y}\right)_0 (y - y_0)}_{\text{height of tangent plane}}$$

The function describing the graph of the tangent plane is often called the *linearisation* of $f(x.y)$ at (x_0, y_0)—it is the linear function that gives the best approximation to $f(x, y)$ for values (x, y) close to (x_0, y_0).

An eqivalent form of the approximation is obtained by using Δ notation: if we set

$$\Delta x = x - x_0, \quad \Delta y = y - y_0, \quad \Delta w = w - w_0$$

we get

$$\Delta w \approx \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y, \quad \text{if } \Delta x \approx 0, \Delta y \approx 0$$

The analagous approximation formula for a function $w = f(x, y, z)$ of three variables would be

$$\Delta w \approx \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y \left(\frac{\partial w}{\partial z}\right)_0 \Delta z, \quad \text{if } \Delta x \approx 0, \Delta y \approx 0, \Delta z \approx 0$$

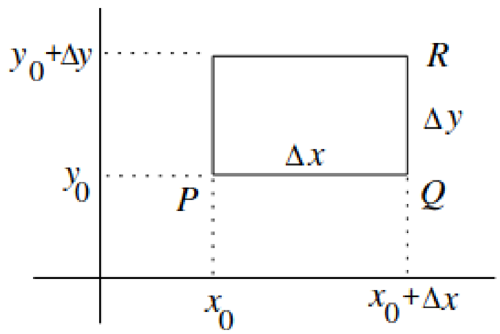
(Note however that this isn't a valid proof, since for functions of three or more variables, we can't use a geometric argument for the approximation formula)

B.3.4 A non-geometrical argument for the approximation formula

We wish to justify—without using reasoning based on 3-space—the approximation formula

$$\Delta w \approx \left(\frac{\partial w}{\partial x}\right)_0 \Delta x + \left(\frac{\partial w}{\partial y}\right)_0 \Delta y, \quad \text{if } \Delta x \approx 0, \Delta y \approx 0$$

Consider the (topographical) illustration:



We are trying to calculate the change in w as we go from P to R in the picture, where $P = (x_0, y_0), R = (x_0 + \Delta x, y_0 + \Delta y)$. This change can be thought of taking place in two steps:

$$\Delta w = \Delta w_1 + \Delta w_2$$

the first being the change in w moving from P to Q, and the second from Q to R; now see that

$$\begin{aligned}
\Delta w_1 &\approx \frac{d}{dx} f(x, y_0)\bigg|_{x_0} \cdot \Delta x = f_x(x_0, y_0)\Delta x \\
&= \left(\frac{\partial w}{\partial x}\right)_0 \Delta x, \\
\Delta w_2 &\approx \frac{d}{dy} f(x_0 + \Delta x, y)\bigg|_{y_0} \cdot \Delta y = f_y(x_0 + \Delta x, y_0)\Delta y \\
&\approx f_y(x_0, y_0)\Delta y = \left(\frac{\partial w}{\partial y}\right)_0 \Delta y
\end{aligned}$$

assuming continuity, substituting the two approximate values gives us the approximation formula. (one can form a proof of by replacing the approximations with equalities using the mean value theorem).

This argument readily generalises to higher-dimensional approximation formulas; the continuity in the neighbourhood of (x_0, y_0, z_0) must, however, be fufilled.

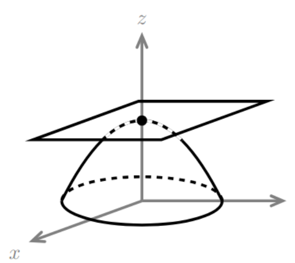
B.3.5 Optimisation Problems

Critical points

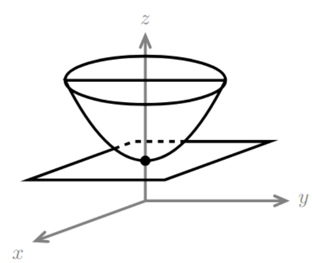
Just as with single variable calculus we consider maxima and minima at points (x_0, y_0) where the first derivatives are 0. As such we define a critical point as any point (x_0, y_0) where

$$\frac{\partial f}{\partial x}(x_0, y_0) = 0 \text{ and } \frac{\partial f}{\partial y}(x_0, y_0) = 0$$

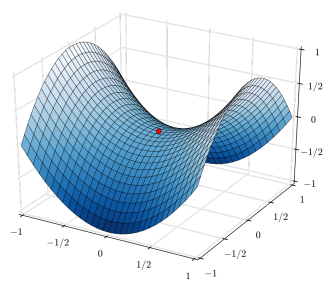
(often abbreviated as $f_x = 0$ and $f_y = 0$) Critical points occur where the tangent plane is horizontal; three types of points (maximum, minimum, saddle) can occur:



Max. with horizontal tang. plane



Min. with horizontal tang. plane


Saddle point

This can also be seen from the formula for the tangent plane; since horizontal plans are of the form $z =$constant, and the equation of the tangent plane at (x_0, y_0, z_0) is

$$z = z_0 + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

we see that it is horizontal when

$$\frac{\partial f}{\partial x}(x_0, y_0) = 0 \text{ and } \frac{\partial f}{\partial y}(x_0, y_0) = 0$$

B.3.6 Least Squares Interpolation

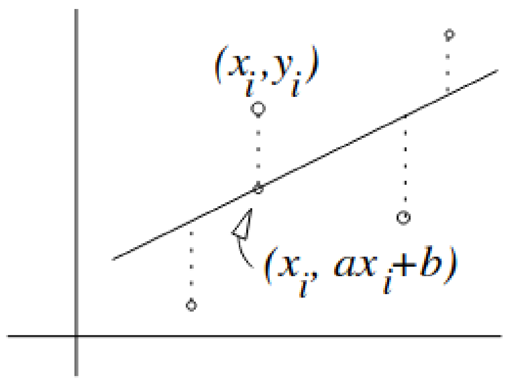
Least Squares interpolation can be framed as a easily solvable multivariate optimisation problem. Consider finding the 'best fit' line through a set of data points

$$(x_1, y_1), (x_2.y_2), \ldots, (x_n, y_n)$$

say we want a line of the equation

$$y = ax + b$$

(Assuming our errors are normally distributed) A good choice for a and b might be the one which minimises the sum of the squared error (squared to consider absolute error);



this means the sum D:

$$D = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

is a *minimum* (over the choice of parameters a and b).

This prescription for finding the line is caled the *method of least squares*, and the resulting line is called the least-squares or the *regression* line.

(next page)

We had

$$D = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

and wanted to minimise D; we see therefore that the the partial derivatives with respect to a and b are 0:

$$\frac{\partial D}{\partial a} = \sum_{i=1}^{n} 2(y_i - ax_i - b)(-x_i) = 0$$
$$\frac{\partial D}{\partial b} = \sum_{i=1}^{n} 2(y_i - ax_i - b)(-1) = 0$$

simplifying, we get

$$\left(\sum x_i^2\right) a + \left(\sum x_i\right) b = \sum x_i y_i$$
$$\left(\sum x_i\right) a + nb = \sum y_i$$

These equations can be divided by n to make them more expressive

$$\bar{s}a + \bar{x}b = \frac{1}{n} \sum x_i y_i$$
$$\bar{x}a + b = \bar{y}$$

where $\bar{x}$ and $\bar{y}$ are the averages of x_i and y_i and $\bar{s} = \sum x_i^2/n$ is the average of the squares. (notice how the equations form a 2×2 system of linear equations; we can use linear algebra to determine a and b)

Fitting curves by least squares

In some cases it might make more sense to fit a polynomial:

$$y = a_0 + a_1 x + a_2 x^2$$

Where we have more parameters to optimise over (in this case three). We once again seek the values of a_0, a_1, a_2 for which the sum of the squares of the error:

$$D = \sum_{i=1}^{n} (y_i - (a_0 + a_1 x_i + a_2 x_i^2))^2$$

is a minimum; we now have three partial derivatives and a square system of three linear equations. Note this is not limited to polynomials; in general, this method of least squares applies to trial expressions of the form

$$y = a_0 f_0(x) + a_1 f_1(x) + \ldots + a_r f_r(x)$$

where we have linear combinations of the functions $f_i(x)$.

B.3.7 Second Derivative Test I

The idea behind a second derivative test is to determine the nature of a critical point; we have three possible outcomes—maximum, minimum, or saddle.

Intuition using general quadratic case
First consider the formula for the roots of a general quadratic formula (can be derived by completing the square); in general two this appears in one of two forms:

$$Ax^2 + Bx + C = 0 \implies x = \frac{-B^2 \pm \sqrt{B^2 - 4AC}}{2A}$$
$$Ax^2 + 2Bx + C = 0 \implies x = \frac{-B^2 \pm \sqrt{B^2 - AC}}{2A}$$

Consider that for the quadratic equation $Ax^2 + 2Bx + C$, we have the following lemma:

$$\begin{aligned}
AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 &\implies Ax^2 + 2Bx + C > 0 \quad \text{for all } x; \\
AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 &\implies Ax^2 + 2Bx + C < 0 \quad \text{for all } x; \\
AC - B^2 < 0, \quad &\implies \begin{cases} Ax^2 + 2Bx + C > 0, & \text{for some } x; \\ Ax^2 + 2Bx + C < 0, & \text{for some } x \end{cases}
\end{aligned}$$

Proof of Lemma: For the first implication, we note from the quadratic formula in the second form that the zeros of $Ax^2 + 2Bx + C$ are imaginary—there are no real zeros. Therefore the (quadratic) graph must lie entirely on one side of the x-axis; the specific side in question can be determined from either A or C, since

$$\begin{aligned}
A > 0 &\implies \lim_{x \to \infty} Ax^2 + 2Bx + C = \infty; \\
C > 0 &\implies Ax^2 + 2Bx + C > 0 \text{ when } x = 0
\end{aligned}$$

(since the quadratic curve has no real roots and lies above the curve it always lies above the curve)

If $A < 0$ or $C < 0$, the reasoning is analagous and proves the second implication.

For the final case, if $AC - B^2 < 0$, the quadratic function has two real roots, so its parabolic graph crosses the x-axis twice, and hence lies partly above and partly below it.
(next page)

Proof of the Second-derivative Test in a special case

Linear functions like $w = w_0 + ax + by$ do not, in general, have maximum or minimum points and its second derivatives are all zero. The simplest functions functions to have interesting critical points are the quadratic functions,which we write in the form:

$$w = w_0 + ax + by + \frac{1}{2}(Ax^2 + 2Bxy + Cy^2)$$

Such a function has in general a unique critical point, which we will assume is $(0,0)$; in this case we can determine the partial derivatives:

$$(w_x)_0 = a, \quad (w_y)_0 = b$$
$$w_{xx} = A, \quad w_{xy} = B, \quad w_{yy} = C$$

(the neat outcome here justifies the $\frac{1}{2}$ and $2B$ in the quadratic function) Since $(0,0)$ is a critical point, we have $a = 0$ and $b = 0$ and our quadratic function has the form

$$w - w_0 = \frac{1}{2}(Ax^2 + 2Bxy + Cy^2)$$

We move w_0 to the left side since the tangent plane at $(0,0)$ is the horizontal plane $w = w_0$, and we are interested in whether the graph of the quadratic function lies above or below this tangent plane; this means whether $w - w_0 > 0$ or $w - w_0 < 0$ at points other than the origin. (remember only one critical point exists)

See that for $(x,y) \neq (0,0)$, we can write the function as

$$w - w_0 = \frac{y^2}{2}\left[A\left(\frac{x}{y}\right)^2 + 2B\left(\frac{x}{y}\right) + C\right]$$

Since $y^2 > 0$ if $y \neq 0$; applying our previous lemma to the second factor we have

$$\begin{aligned}
AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 &\implies w - w_0 > 0 \quad \text{for all } (x,y) \neq (0,0);\\
&\implies (0,0) \text{ is a minimum point;}\\
AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 &\implies w - w_0 < 0 \quad \text{for all } (x,y) \neq (0,0);\\
&\implies (0,0) \text{ is a maximum point;}\\
AC - B^2 < 0, \quad &\implies \begin{cases} w - w_0 > 0, & \text{for some } (x,y);\\ w - w_0 < 0, & \text{for some } (x,y);\end{cases}\\
&\implies (0,0) \text{ is a saddle point}
\end{aligned}$$

(we fix $(0,0)$ to be a critical point, and we see how all other points compare to that critical point to decide its nature)
(next page)

Argument/Intuition for the Second-derivative Test for a general function

Consider now a general formula $w = f(x, y)$, and assume it has a critical point at (x_0, y_0) and continuous second derivatives in the neighbourhood of the critical point. Then by a generalisation of Taylor's formula to functions of several variables, the function has a best quadratic approximation at the critical point.

Consider moving the critical point to the origin by making the change of variables

$$u = x - x_0, \quad v = y - y_0$$

Then we have the quadratic approximation

$$w = f(x, y) \approx w_0 + \frac{1}{2}(Au^2 + 2Buv + Cv^2)$$

where the coefficients A, B, C are found as the second partial derivatives with respect to u and v at $(0, 0)$ or the second partial deriatives with respect to x and y at (x_0, y_0). (Intuitively see that the coefficients are found by partial derivation)

It is reasonable to suppose that the quadratic approximation well approximates the function near (x_0, y_0). Using this we can therefore interpret the nature of a critical point by approximation.

B.3.8 Second Derivative Test II

We have a quadratic approximation for a general function $w = f(x, y)$ with critical point $(x_0.y_0)$, where substituting

$$u = x - x_0, \quad v = y - y_0$$

we have

$$\begin{aligned} w &\approx w_0 + w_x u + w_y v + \frac{1}{2}(Au^2 + 2Buv + Cv^2) \\ &\approx w_0 + \frac{1}{2}(Au^2 + 2Buv + Cv^2) \end{aligned}$$

near the critical point (since the first partial derivatives are zero at the critical point)

Showing the significance of the second derivative
Using $u = x - x_0$ and $v = y - y_0$ we can apply the chain rule to show

$$w_x = w_u u_x + w_v v_x = w_u \quad \text{since } u_x = 1 \text{ and } v_x = 0$$

and

$$w_y = w_u u_y + w_v v_y = w_v \quad \text{since } u_y = 0 \text{ and } v_y = 1$$

Therefore at the corresponding points,

$$(w_x)_{(x_0,y_0)} = (w_u)_{(0,0)}, \quad (w_y)_{(x_0,y_0)} = (w_v)_{(0,0)}$$

(this can be used to intuit the first derivative in the approximation formula but here we care about the second derivatives) differentiating once more with the same reasoning we have

$$\begin{aligned} (w_{xx})_{(x_0,y_0)} &= (w_{uu})_{(0,0)} & (w_{xy})_{(x_0,y_0)} &= (w_{uv})_{(0,0)} & (w_{yy})_{(x_0,y_0)} &= (w_{vv})_{(0,0)} \\ &= A & &= B & &= C \end{aligned}$$

all this is to say that we can reduce a general formula into a special quadratic case. After which we can use the values of A, B, C to determine the nature of the critical point.
(next page)

Procedure for second derivative test
With that the second-derivative test therefore has two steps:

1. Find the critical points by solving the simultaneous equations

$$\begin{cases} f_x(x,y) = 0 \\ f_y(x,y) = 0 \end{cases}$$

2. To test the point to see if it is a local minimum or maximum point, we calculate the three second derivatives at the critical point (x_0, y_0), whose values can be denoted as (using notation $(f)_0 = f(x_0, y_0)$):

$$A = (f_{xx})_0, \quad B = (f_{xy})_0 = (f_{yx})_0, \quad C = (f_{yy})_0$$

and finally we have: *letting* (x_0, y_0) *be a critical point of* $f(x,y)$ *and* A, B, C *be denoted as above, then*

$$\begin{aligned} AC - B^2 > 0, \quad A > 0 \text{ or } C > 0 &\implies (x_0, y_0) \text{ is a minimum point;} \\ AC - B^2 > 0, \quad A < 0 \text{ or } C < 0 &\implies (x_0, y_0) \text{ is a maximum point;} \\ AC - B^2 < 0, \quad &\implies (x_0, y_0) \text{ is a saddle point} \end{aligned}$$

If $AC - B^2 = 0$, the test fails as the point is *degenerate*, more investigaion is required.

B.3.9 Total Differentials and Chain Rule

Total Differential
The *Total differential* of $f(x, y, z)$ is given by

$$dw = w_x dx + w_y dy + w_z dz$$

This can be seen from tangent approximation where

$$\Delta w \approx \left.\frac{\partial w}{\partial x}\right|_O \Delta x + \left.\frac{\partial w}{\partial y}\right|_O \Delta y + \left.\frac{\partial w}{\partial z}\right|_O \Delta z$$

letting $\Delta \to 0$ we get the total differential.

Chain rule
Suppose w is a function of x, y and that x, y are functions of u, v; that is

$$w = f(x, y) \text{ and } x = x(u, v), y = y(u, v)$$

The use of the term chain comes from the idea that to compute w we need to do a chain of computations:

$$(u, v) \to (x, y) \to w$$

we will say that w is a *dependent* variable, u and v are *independent* variables, and x and y are *intermediate* variables.

Since w is a function of x and y it has partial derivatives $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial y}$; but since w is also ultimately a function of u and v we can also compute the partial derivatives $\frac{\partial w}{\partial u}$ and $\frac{\partial w}{\partial v}$. The chain rule relates these derivatives by the formulas:

$$\begin{aligned}
\frac{\partial w}{\partial u} &= \frac{\partial w}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial u} \\
\frac{\partial w}{\partial v} &= \frac{\partial w}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial v}
\end{aligned}$$

Chain rule from tangent approximation
We can come to the chain rule using tangent approximation:

$$\Delta w \approx \left.\frac{\partial w}{\partial x}\right|_O \Delta x + \left.\frac{\partial w}{\partial y}\right|_O \Delta y$$

holding v constant and dividing by Δu we get

$$\frac{\Delta w}{\Delta u} \approx \left.\frac{\partial w}{\partial x}\right|_O \frac{\Delta x}{\Delta u} + \left.\frac{\partial w}{\partial y}\right|_O \frac{\Delta y}{\Delta u}$$

letting $\Delta u \to 0$ gives the chain rule for $\frac{\partial w}{\partial u}$.

B.3.10 Gradient

Definition
Given $w = f(x, y)$, the partial derivatives of w can be put together in a vector called the *gradient* of w

$$\operatorname{grad} w = \left\langle \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y} \right\rangle \quad \text{or} \quad \nabla w = \left\langle \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y} \right\rangle$$

Specifying a point $P_0 = (x_0, y_0)$, we can evaluate the gradient at that point; this is given by several notations:

$$\operatorname{grad} w(x_0, y_0) = \nabla w|_{P_0} = \nabla w|_O = \left\langle \left.\frac{\partial w}{\partial x}\right|_O, \left.\frac{\partial w}{\partial y}\right|_O \right\rangle$$

Analogously, for functions $w = f(x, y, z)$, we have the gradient

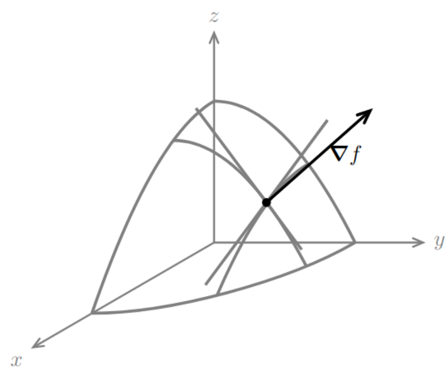
$$\operatorname{grad} w = \nabla w = \left\langle \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}, \frac{\partial w}{\partial z} \right\rangle$$

The gradient takes a *scalar* function $f(x, y)$ and produces a vector ∇f. The gradient has many geometric properties; one of them is that the gradient is perpendicular to the level curves $f(x, y) = c$.
(newpage)

Proof that the gradient is perpendicular to the level curves and surfaces

Letting $w = f(x, y, z)$ be a function of 3 variables. Here we show that at any point $P = (x_0, y_0, z_0)$ on the level surface $f(x, y, z) = c$ (so $f(x_0, y_0, z_0) = c$) the gradient $\nabla f|_P$ is perpendicular to the surface.

By this we mean it is perpendicular to the tangent to any curve that lies on the surface and goes through P:



Consider the position vector

$$\boldsymbol{r}(t) = \langle x(t), y(t), z(t) \rangle$$

be a curve on a level surface with $\boldsymbol{r}(t_0) = \langle x_0, y_0, z_0 \rangle$. We let $g(t) = f(x(t), y(t), z(t)) = c$ (a level curve). Differentiating the equation of the level surface with respect to t, the chain rule gives

$$\frac{dg}{dt} = \left.\frac{\partial f}{\partial x}\right|_P \left.\frac{dx}{dt}\right|_{t_0} + \left.\frac{\partial f}{\partial y}\right|_P \left.\frac{dy}{dt}\right|_{t_0} + \left.\frac{\partial f}{\partial z}\right|_P \left.\frac{dz}{dt}\right|_{t_0} = 0$$

since $g(t)$ is a fixed scalar. In vector form this gives

$$\left\langle \left.\frac{\partial f}{\partial x}\right|_P, \left.\frac{\partial f}{\partial y}\right|_P, \left.\frac{\partial f}{\partial z}\right|_P \right\rangle \cdot \left\langle \left.\frac{dx}{dt}\right|_{t_0}, \left.\frac{dy}{dt}\right|_{t_0}, \left.\frac{dz}{dt}\right|_{t_0} \right\rangle = 0$$
$$\implies \nabla f|_P \cdot \boldsymbol{r}'(t_0) = 0$$

Since the dot product is 0, we have shown that the gradient is perpendicular to the tangent of any curve that lies on the level surface. (see how this applies to any level surface)

B.3.11