QMSS Math Camp

Calculus/Analysis

Emile Esmaili

ede2110@columbia.edu

Outline

- Warmup
- Limits
- Differential calculus (single and multivariable)
- Optimization
- Sequences and Series
- Integration

Warmup

Math Basics

- **The natural numbers**, $\mathbb{N}$, are $1, 2, 3, \dots$ and allow us to count.
- The **integer numbers**, $\mathbb{Z}$, include the natural numbers (positive integers), their negative counterparts, and 0: $\dots, -2, -1, 0, 1, 2, \dots$
- The **rational numbers**, $\mathbb{Q}$, consist of all numbers that can be written as a ratio of two integers, $\frac{n}{m}$, with $m \neq 0$. For example, $-\frac{1}{2}$ and $\frac{123}{4}$
- The **real numbers**, $\mathbb{R}$, include all of the rational numbers along with the irrational numbers, such as $\sqrt{2} \approx 1.41421$ or $e \approx 2.71828$, or π.
- The **complex numbers**, $\mathbb{C}$, are of the form $a + ib$, where $a, b \in \mathbb{R}$ and where $i^2 = -1$. In the complex numbers, we can solve any polynomial equation. We note $\mathfrak{Re}(z) = a$ and $\mathfrak{Im}(z) = b$

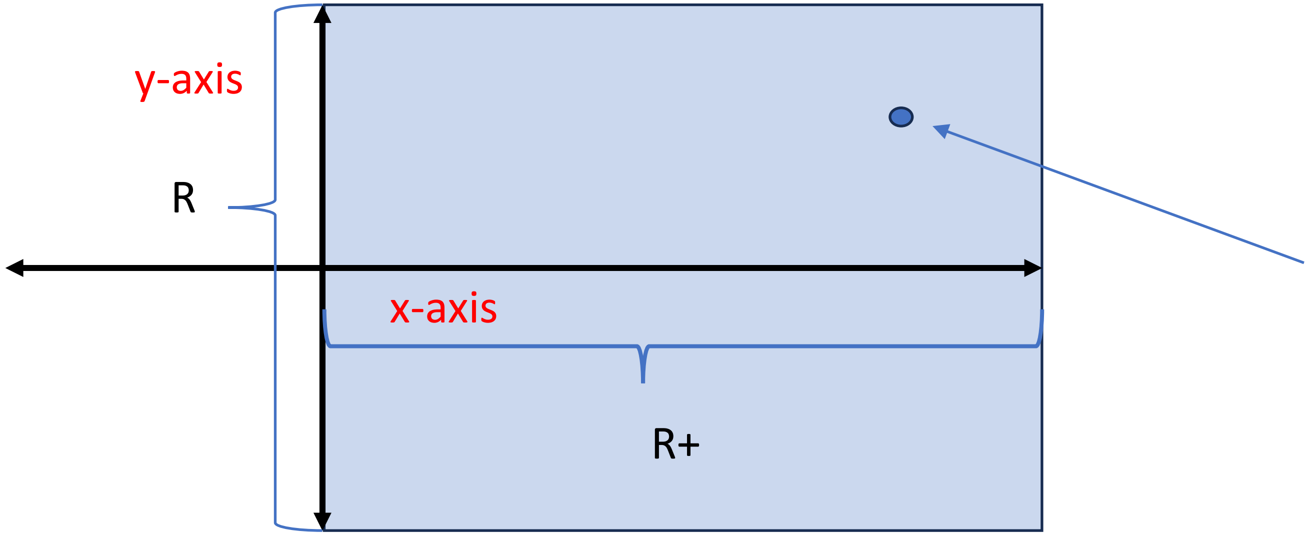
Remember that $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

Math basics

- "In" notation: $a \in A$, where a is an element in the set A.
- "For all" notation: $\forall x \in S$, where it means "for all x in the set S."
- "There exists" notation: $\exists x \in S$, where it means "there exists an x in the set S."
- "R+" notation: $\mathbb{R}^+$, where it represents the set of positive real numbers.
- "R*" notation: $\mathbb{R}^*$, where it represents the set of non-zero real numbers.
- Set inclusion notation: $A \subseteq B$, where it means "set A is a subset of set B."
- Set exclusion notation: $A \backslash B$, where it means "set A excluding the elements in set B."
- a closed interval contains its frontier points and is noted [a,b]
- an open interval does not contain its frontier points and it noted (a,b) or]a,b[

Math Basics

The Cartesian product of two elements in sets A and B is denoted as $A \times B$
For instance $(x, y) \in \{\mathbb{R}^+ \times \mathbb{R}\}$ means $x \in \mathbb{R}^+$ and $y \in \mathbb{R}$



One point (x,y) in our set

We note $\mathbb{R} \times \mathbb{R}$ as $\mathbb{R}^2$

Polynomials

Definition: We note $P(x)$ a polynomial in x:

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

where:

- $P(x)$ is the polynomial function.
- $a_n, a_{n-1}, \ldots, a_2, a_1, a_0$ are coefficients.
- x is the variable.
- n is a non-negative integer and represents the highest degree of the polynomial.

Example: The quadratic polynomial is a second-degree polynomial and can be written as:

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

$Q(x) = 2x^2 - 3x + 1$ is a second-degree polynomial.

Polynomials exercise

Given Expressions:
$P(X) = X^3 + 3X^2 - 1,\ Q(X) = -X^3 - X + 1,$
Calculate $(P + Q)(X)$:

$$\begin{aligned}(P+Q)(X) &= P(X) + Q(X) \\ &= X^3 + 3X^2 - 1 + (-X^3 - X + 1) \\ &= (X^3 - X^3) + 3X^2 - X + 1 - 1 \\ &= 3X^2 - X.\end{aligned}$$

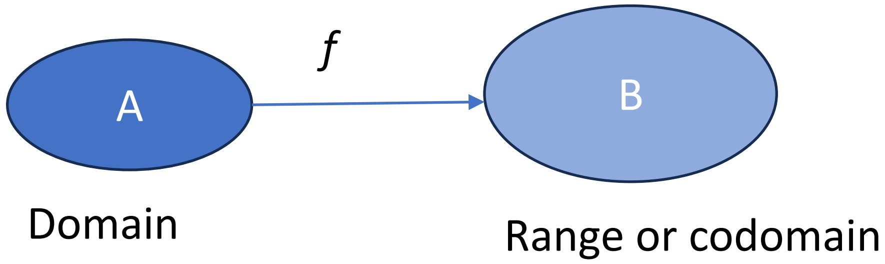
Given Expressions:
$P(X) = X^2 + X + 1,\ Q(X) = -X + 1,$
Calculate $(PQ)(X)$:

$$\begin{aligned}(PQ)(X) = P(X)Q(X) &= (X^2 + X + 1)(-X + 1) \\ &= -X^3 + X^2 - X^2 + X - X + 1 \\ &= -X^3 + 1.\end{aligned}$$

Given Expressions:
$P(X) = X^2 + X + 1,\ Q(X) = X^2 + 1,$
Calculate $(P(Q))(X)$:

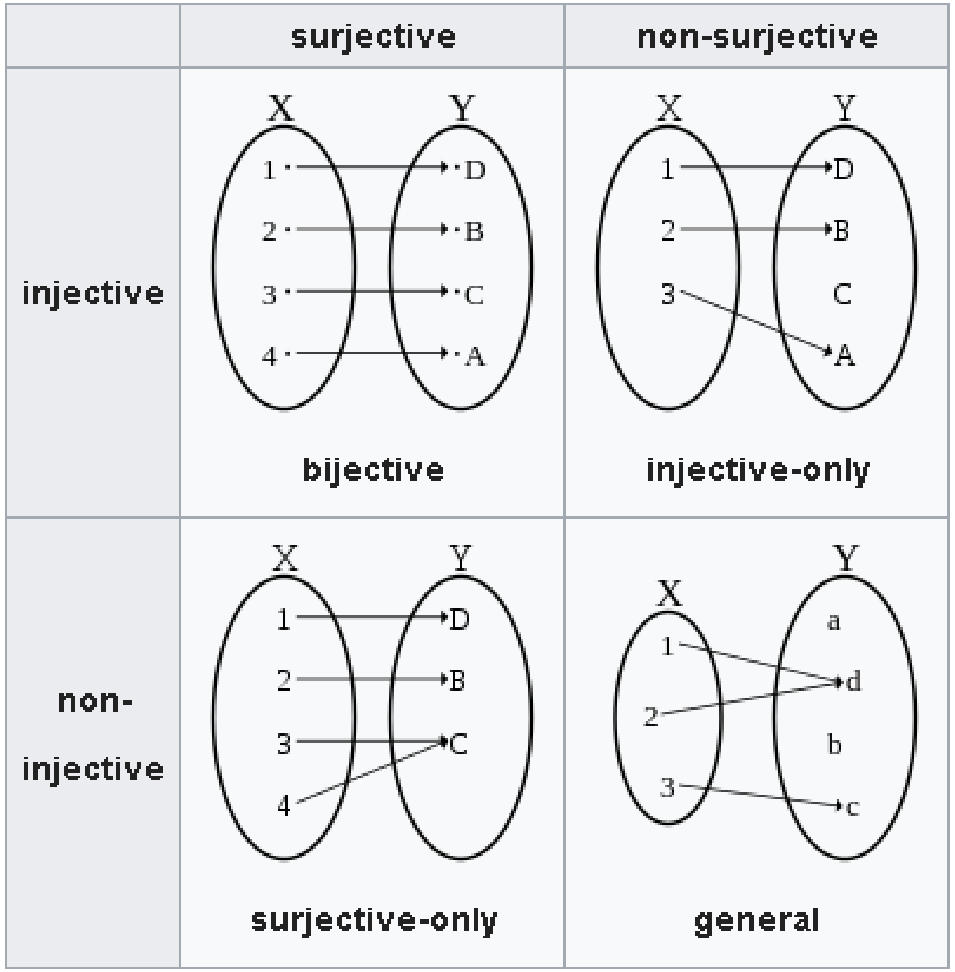
$$\begin{aligned}(P(Q))(X) &= (Q(X))^2 + Q(X) + 1 \\ &= (X^2 + 1)^2 + (X^2 + 1) + 1 \\ &= X^4 + 2X^2 + 1 + X^2 + 1 + 1 \\ &= X^4 + 3X^2 + 3.\end{aligned}$$

functions



- **Function:** A function $f : A \to B$ is a rule that assigns to each element $a \in A$ a unique element $b \in B$.
- **Injective (One-to-One):** A function f is said to be injective if it maps distinct elements in the domain A to distinct elements in the codomain B. In other words, $\forall a_1, a_2 \in A,\ f(a_1) = f(a_2) \iff a_1 = a_2$.
- **Surjective (Onto):** A function f is said to be surjective if, $\forall b \in B, \exists a \in A$ such that $f(a) = b$. In other words, the range of f covers the entire codomain B.
- **Bijective:** A function f is said to be bijective if it is both injective and surjective. It means that f is a one-to-one correspondence between the elements of A and B.

functions



C has no inverse image, so not surjective

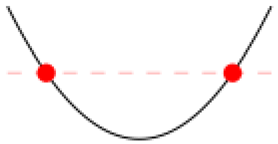
F(3) = F(4), but
3 != 4

functions

The function $f(x) = x^2$, considered from $\mathbb{R} \to \mathbb{R}$

- **Not Surjective :** there is no real number x such that $x^2 = -1$. Therefore, $f(x) = x^2$ is not surjective in codomain $\mathbb{R}$
- **Not Injective :** both $x = 2$ and $x = -2$ result in $f(x) = 4$, so it fails the one-to-one property.

In summary, $f(x) = x^2$ is neither surjective nor injective when considered over the real numbers.



Not injective

Non monotonous

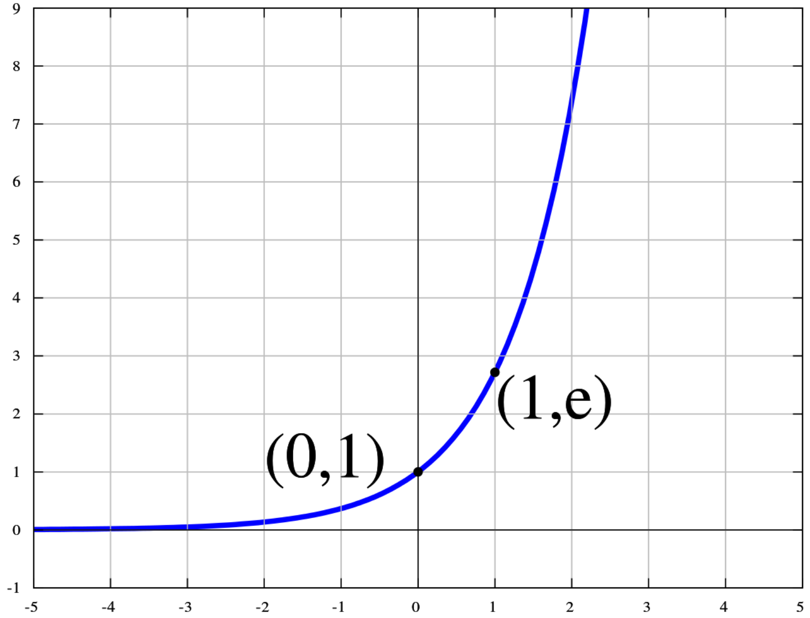
injective

monotonous

exponential

An exponential function is a function of the form: $f(x) = a^x$ where $a > 0$.

- The most common exponential function is: $y = \exp(x) = e^x$
- **Product Rule:** $e^{x_1} \cdot e^{x_2} = e^{x_1 + x_2}$
- **Quotient Rule:** $\frac{e^{x_1}}{e^{x_2}} = e^{x_1 - x_2}$
- **Power Rule:** $(e^x)^a = e^{x \cdot a}$

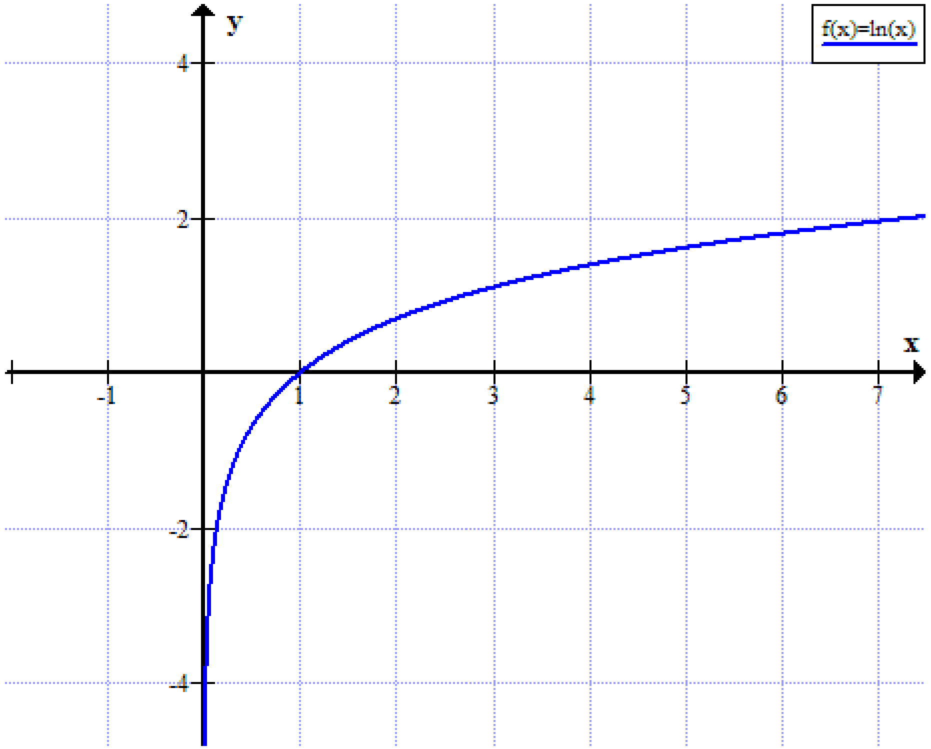


Logarithm

the Logarithm function is noted $log(x)$

- $\log_b(x) = y \iff x = b^y$.
- Logarithms to base e are called natural logarithms: $\ln(x)$.
- **Product Rule:** $\log(x_1 \cdot x_2) = \log(x_1) + \log(x_2)$
- **Quotient Rule:** $\log\left(\frac{x_1}{x_2}\right) = \log(x_1) - \log(x_2)$
- **Power Rule:** $\log(x^a) = a \cdot \log(x)$

WARNING
$ln(x)$ is defined: $\mathbb{R}^{+*} \to \mathbb{R}$
$\ln(0)$ does not exist



Logarithm and exponential are inverse

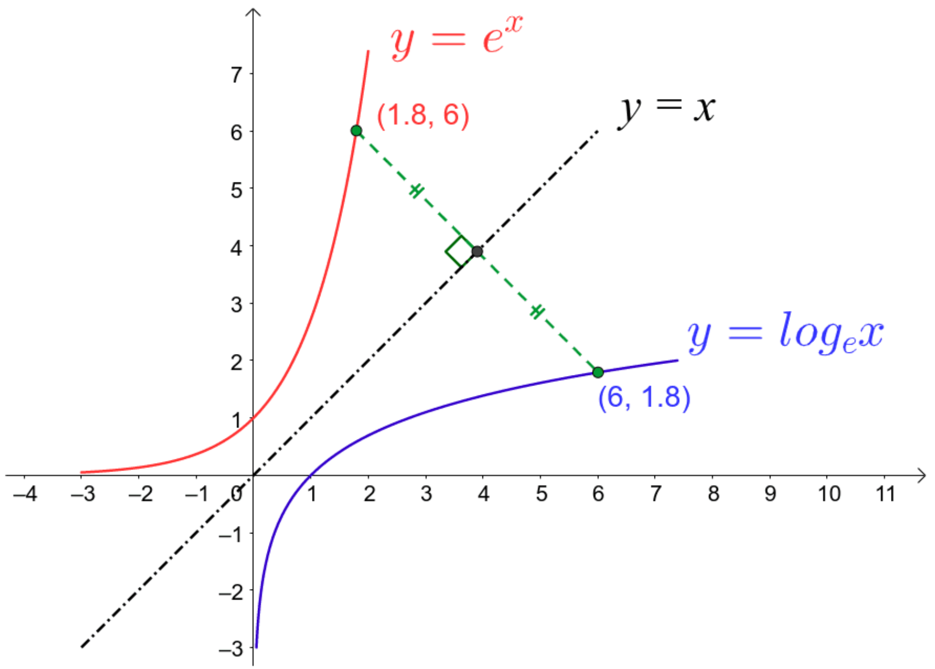
Let f be a function from set A to set B. If there exists a function f^{-1} from set B to set A such that for all x in A and y in B, the following holds:

$$f^{-1}(f(x)) = x \text{ for all } x \text{ in } A$$

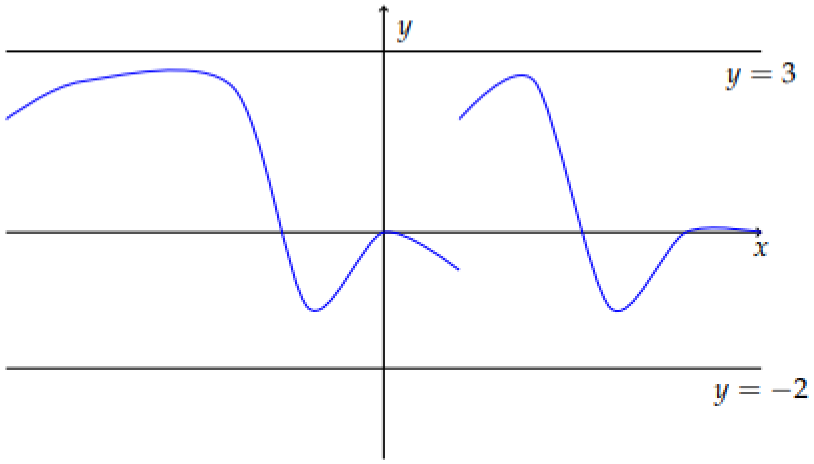
$$f(f^{-1}(y)) = y \text{ for all } y \text{ in } B$$

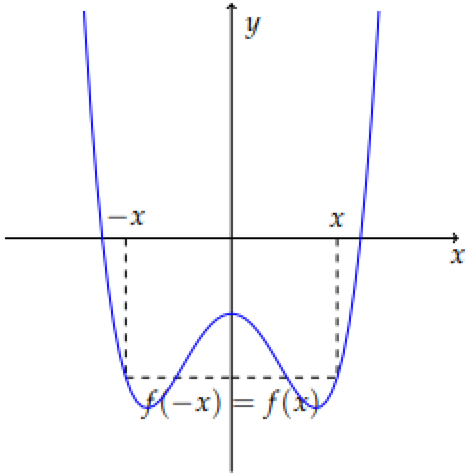
then f and f^{-1} are inverse functions.

- $\log_a(a^x) = x$; $a^{\log_a(x)} = x$.
- In particular,
  - $\ln(e^x) = \log_e(e^x) = x$
  - $e^{\ln(x)} = e^{\log_e(x)} = x$

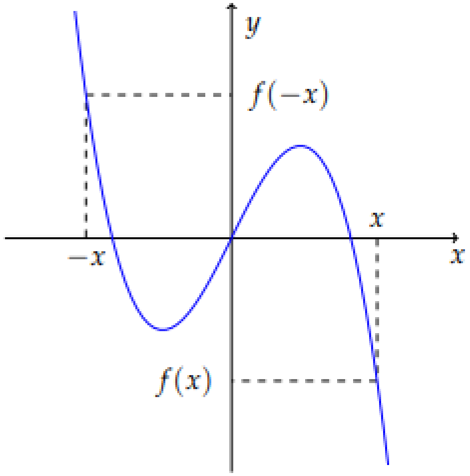


Functions – additional vocab





Even
f(-x) = f(x)



Odd
f(-x) = -f(x)

An **upper bound** of f is 3
A **lower bound** of f is -2

Definition: Bounded Function

A function $f : A \to \mathbb{R}$ is said to be bounded if $\exists\ M \in \mathbb{R}$ such that $\forall\ x \in A$, we have $|f(x)| \leq M$.

Functions

The absolute value of a real number x, denoted as $|x|$, is defined as follows:

$$|x| = \begin{cases} x, & \text{if } x \geq 0 \\ -x, & \text{if } x < 0 \end{cases}$$

For all $x, y \in \mathbb{R}$:

1. $|xy| = |x| \cdot |y|$
2. $|x + y| \leq |x| + |y|$ (Triangle Inequality)
3. $|x + y| \geq ||x| - |y||$ (Reverse Triangle Inequality)

Limits

$$lim_{x\to\infty}\frac{1}{x} = \frac{1}{\text{Big number}} = 0$$

$$lim_{x\to 0}\frac{1}{x} = \frac{1}{\text{small number}} = \infty$$

How to rigorously formalize this?

Limits

$$\forall \epsilon > 0, \exists \delta > 0 : \forall x \text{ in Dom(f)} \;,\; \text{if } 0 < |x - a| < \delta, \;\text{then } |f(x) - L| < \epsilon.$$

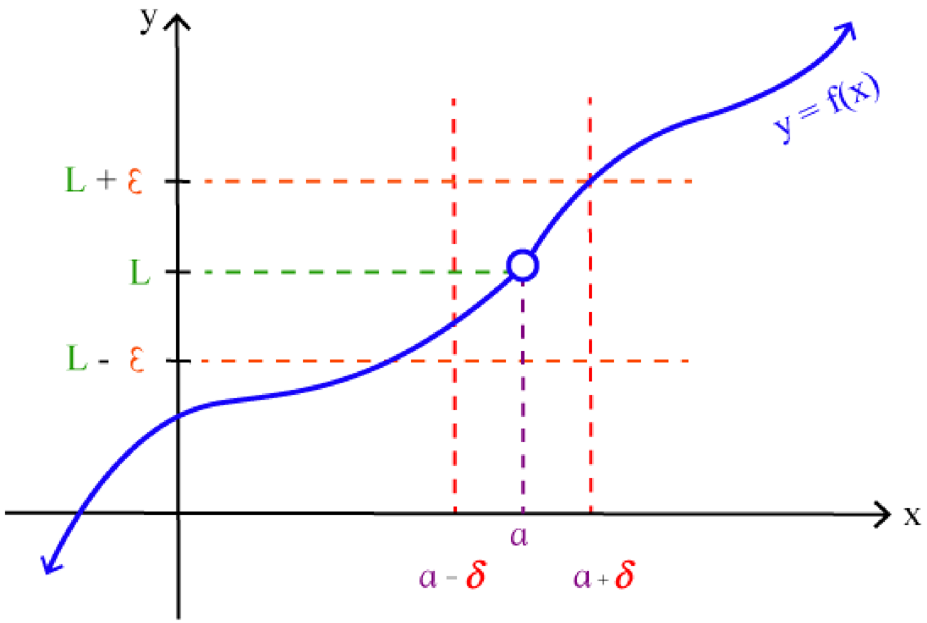
Let $f(x)$ be a function defined on the interval that contains $x = a$.

Then $\lim_{x \to a} f(x) = L$ if for every number $\varepsilon > 0$ there exists some real number $\delta > 0$ so that if

$$0 < |x - a| < \delta \text{ then } |f(x) - L| < \varepsilon$$

As ***x*** goes near a (within delta)

f cannot escape **L**
(even if **epsilon** is 0.000001)



Limits

Indeterminate Forms in Limits:

- $\frac{0}{0}$ - Zero divided by zero.
- $\frac{\infty}{\infty}$ - Infinity divided by infinity.
- $0 \cdot \infty$ - Zero times infinity.
- $\infty - \infty$ - Infinity minus infinity.

Useful limits

For any positive integer $n > 0$, the following limits hold:

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

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

$$\lim_{x \to +\infty} \frac{\ln(x)}{x^n} = 0$$

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

Limits quick workout

1.

$$lim_{x\to\infty}\frac{4x^2+3x+1}{2x^4+1}$$

Solution: The limit at positive or negative infinity of a quotient of polynomials is the limit of the terms with the highest degree. To find it, factorize the expression:

$$\frac{4x^2+3x+1}{2x^4+1}=\frac{4x^2}{2x^4}\cdot\frac{1+\frac{3}{4x}+\frac{1}{2x^2}}{1+\frac{1}{2x^4}}$$

Simplifying further:

$$\frac{2}{x^2}\cdot\frac{1+\frac{3}{4x}+\frac{1}{2x^2}}{1+\frac{1}{2x^4}}$$

The second fraction approaches 1 as x tends to infinity, and the first fraction approaches 0. Therefore, the requested limit is 0.

2.

$$\lim_{x\to+\infty}\frac{1}{\sqrt{x+1}-\sqrt{x-1}}$$

Solution: We cannot determine the limit from this form; it's an indeterminate form. We multiply by the conjugate quantity. For $x\geq 1$, we have:

$$\frac{1}{\sqrt{x+1}-\sqrt{x-1}}=\frac{\sqrt{x+1}+\sqrt{x-1}}{(\sqrt{x+1}+\sqrt{x-1})(\sqrt{x+1}-\sqrt{x-1})}$$

Simplifying further $(a-b)(a+b)=a^2-b^2$:

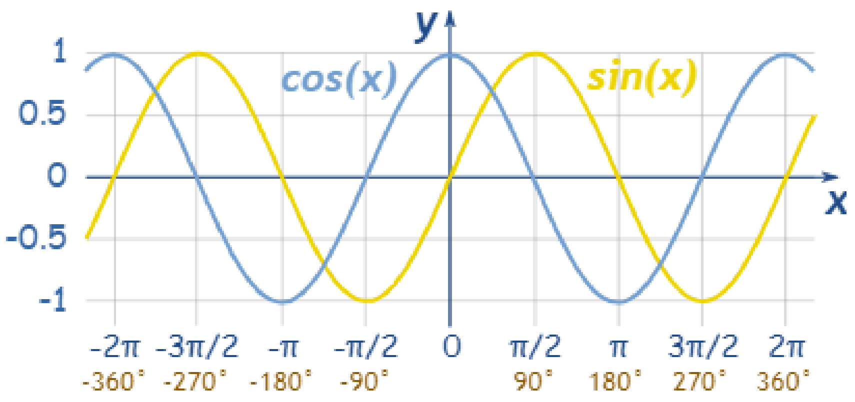
$$\frac{\sqrt{x+1}+\sqrt{x-1}}{(x+1)-(x-1)}=\frac{\sqrt{x+1}+\sqrt{x-1}}{2}$$

In this form, it's clear: the limit is ∞

Trigonometry

The functions sin and cos are 2π-periodic, meaning $\mathbf{cos(x+2k\pi) = cos(x)}$, $k \in \mathbb{Z}$
Moreover:

- The cosine function is even, and the sine function is odd. This means that for all $x \in \mathbb{R}$, $\cos(-x) = \cos(x)$ and $\sin(-x) = -\sin(x)$.
- For all $x \in \mathbb{R}$, $\cos(x+\pi) = -\cos(x)$ and $\sin(x+\pi) = -\sin(x)$.
- For all $x \in \mathbb{R}$, $\cos(x) = \cos(2\pi n + x)$ and $\sin(x) = \sin(2\pi n + x)$, where n is an integer.



Some useful identities

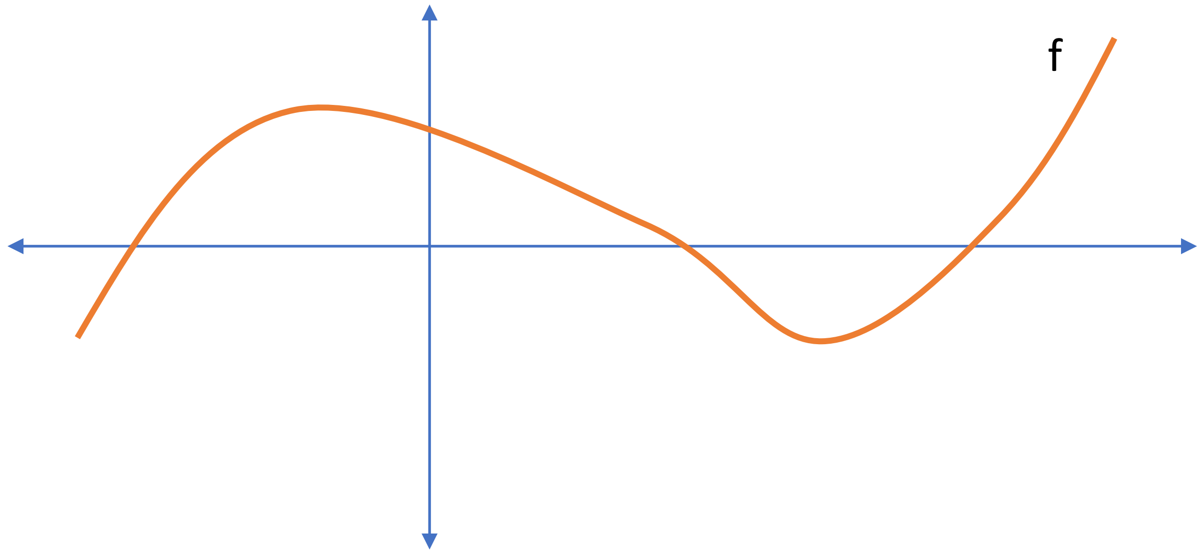
- $cos^2(x) + sin^2(x) = 1$
- $cos(x+y) = cos(x)cos(x) - sin(x)sin(y)$
- $sin(x+y) = sin(x)cos(xy) + cos(x)sin(y)$
- $cos(2x) = 2cos^2(x) - 1$

Differential Calculus

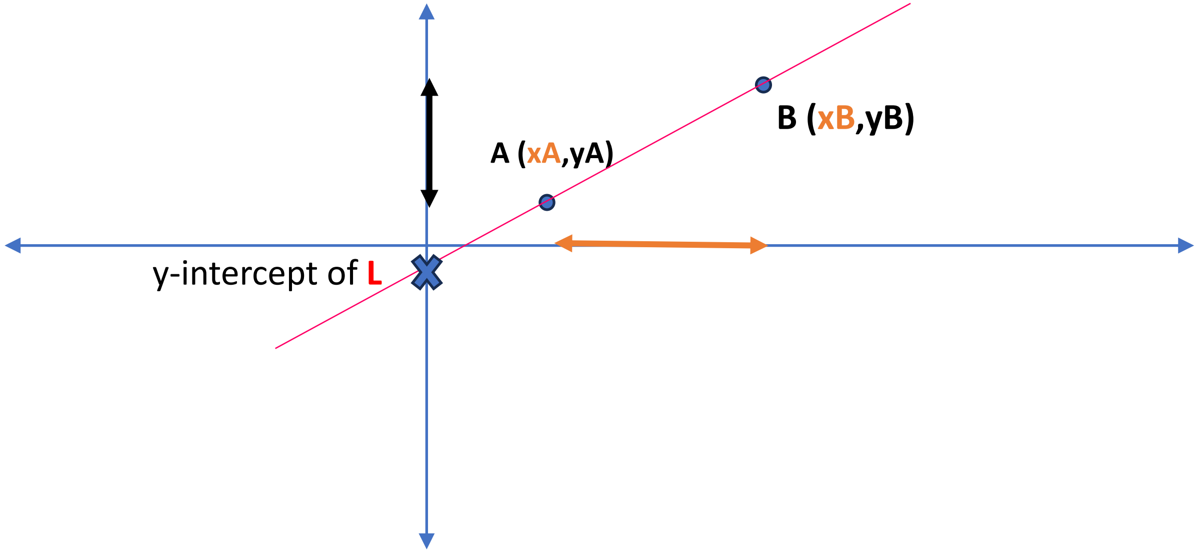
Differential Calculus – Single Variable

$$\mathbb{R} \longrightarrow \mathbb{R}$$

$$x \longrightarrow f(x)$$

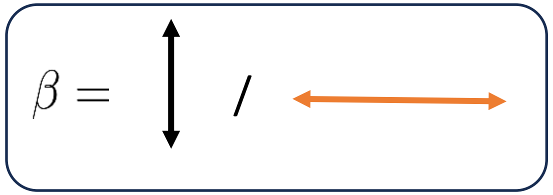


Equation of a line in the plane



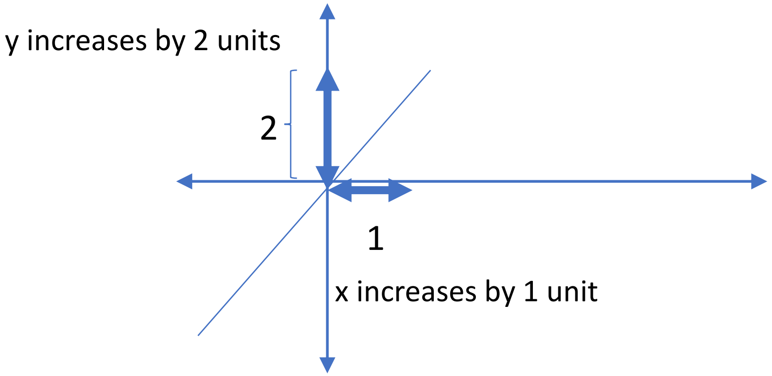
Line **L** between **A** and **B has slope:**

$$\beta = \frac{yB - yA}{xB - xA}$$



Every line in $\mathbb{R}^2$ has equation $y = a + \beta x$ with β the slope, and a the y-intercept

Ex: y = 2x + 0



Differential Calculus – Single Variable

$f : I \to \mathbb{R}$ and $a \in I$.

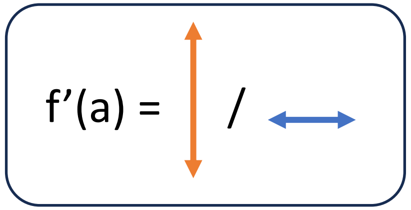
f is differentiable in a if the following limit exists

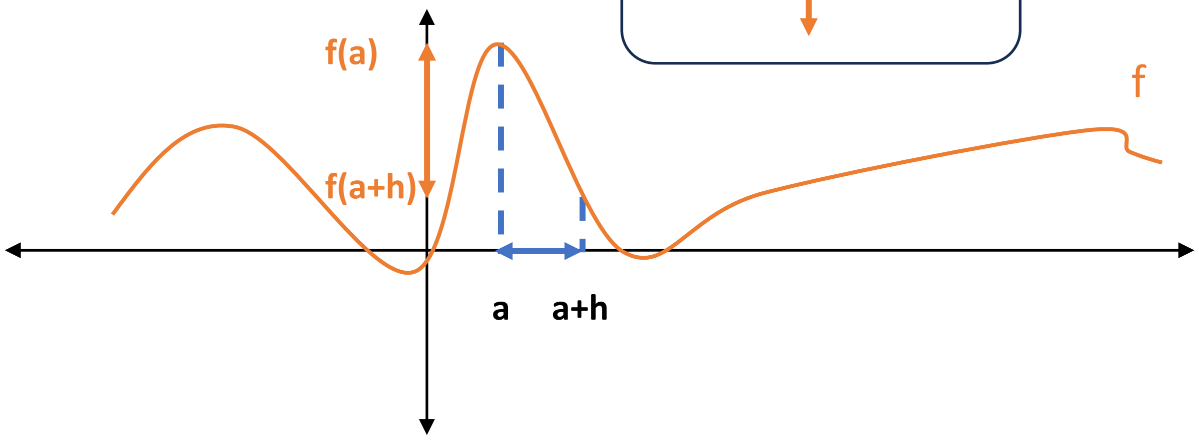
$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

This limit is the derivative of f at point a, noted $f'(a)$.

$f'(a)$ also noted $\frac{df}{da}$

where d(.) notes a small change or *delta* in a variable





the derivative tells you how sensitive the output **f(a)** is to the input **a**

Differential Calculus – Single Variable

Let's try this definition to compute a simple derivative

$$
\begin{aligned}
f(x) &= x^2 \\
f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} \\
&= \lim_{h \to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\
&= \lim_{h \to 0} \frac{2xh + h^2}{h} \\
&= \lim_{h \to 0} \frac{h(2x + h)}{h} \\
&= \lim_{h \to 0} (2x + h) \\
&= 2x + 0 \\
&= 2x
\end{aligned}
$$

Derivatives Toolbox

$f(x) = c$	$f'(x) = 0$
$f(x) = x$	$f'(x) = 1$
$f(x) = x^n$	$f'(x) = nx^{n-1}$
$f(x) = \frac{1}{x}$	$f'(x) = -\frac{1}{x^2}$
$f(x) = \sqrt{x}$	$f'(x) = \frac{1}{2\sqrt{x}}$
$f(x) = x^\alpha$	$f'(x) = \alpha x^{\alpha-1}$
$f(x) = \ln x$	$f'(x) = \frac{1}{x}$
$f(x) = e^x$	$f'(x) = e^x$
$f(x) = \sin x$	$f'(x) = \cos x$
$f(x) = \cos x$	$f'(x) = -\sin x$

Product Rule: If u and v are functions of x, then the derivative of their product is given by

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

Quotient Rule: If u and v are functions of x, then the derivative of their quotient is given by

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2} \quad \text{if } v \neq 0$$

Derivatives Toolbox – Composition

Let I be an interval in $\mathbb{R}$, $f : I \to \mathbb{R}$, and $g : \mathbb{R} \to \mathbb{R}$ be two functions. We define the composition function $g \circ f$, which maps from I to $\mathbb{R}$, as follows:

$$\forall x \in I, (g \circ f)(x) = g(f(x)).$$

Ex:

$$f(x) = 2x + 1 \quad \text{and} \quad g(x) = x^2 + 3.$$

We can then compute the composition function $g \circ f$ as follows:

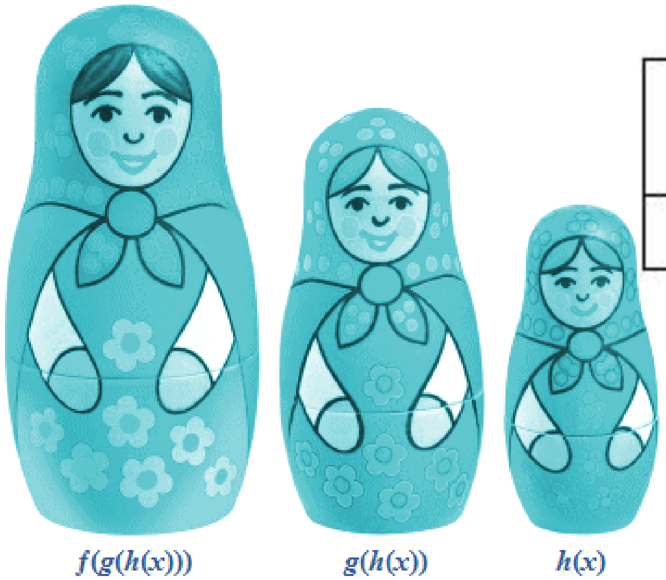
$$(g \circ f)(x) = g(f(x)) = f(x)^2 + 3 = (2x + 1)^2 + 3$$

Derivatives Toolbox

Let $f:(a,b)\to\mathbb{R}$ and $g:\mathbb{R}\to\mathbb{R}$ be two differentiable functions. Then the composite function $g\circ f:(a,b)\to\mathbb{R}$ is also differentiable, and $\forall x\in(a,b)$, its derivative is given by:

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

CHAIN RULE (think of Russian dolls)



$\left(\frac{1}{u}\right)'=-\frac{u'}{u^2}$	$(\sqrt{u})'=\frac{u'}{2\sqrt{u}}$	$(u^\alpha)'=\alpha u'u^{\alpha-1}$	$(\ln u)'=\frac{u'}{u}$
$(e^u)'=u'e^u$	$(\sin u)'=u'\cos u$	$(\cos u)'=-u'\sin u$	...

Derivatives - workout

$f : x \mapsto \sin(3x^2 + \frac{1}{x})$

f is defined and differentiable for all $x \neq 0$. It can be expressed as the composition $f \circ g$ of $f(X) = \sin(X)$ and $g(x) = 3x^2 + \frac{1}{x}$.

Its derivative for all $x \neq 0$ is therefore

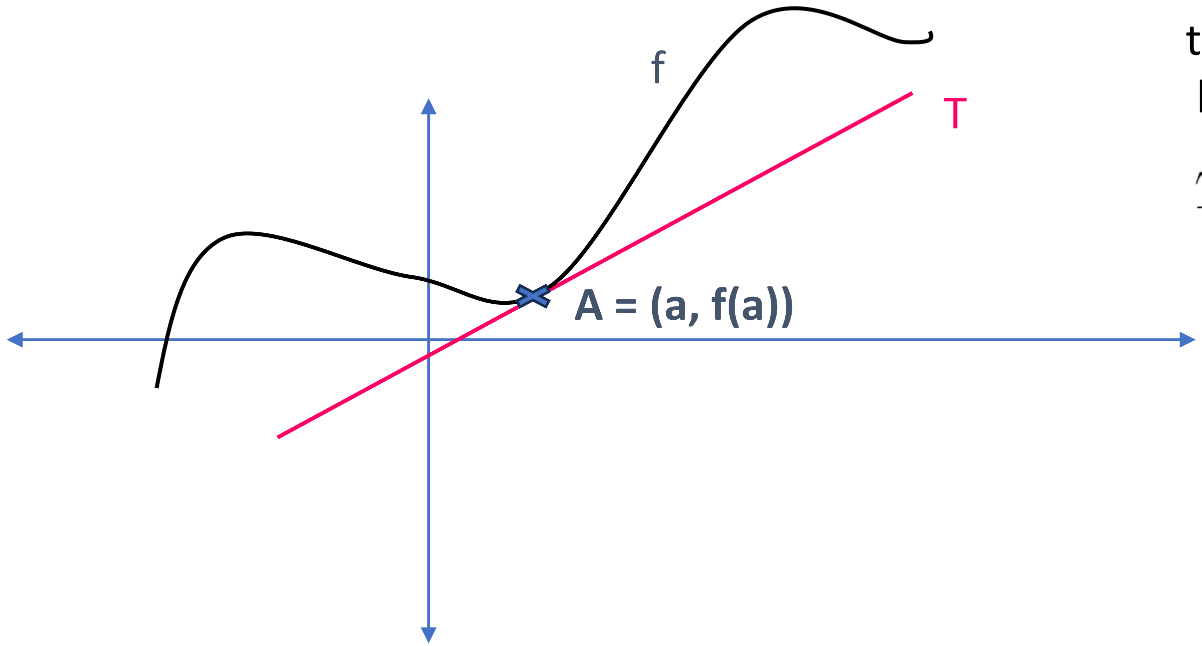
$$\begin{aligned} f'(x) &= \sin'(f(x)) \cdot g'(x) \\ &= \left(6x - \frac{1}{x^2}\right) \cos\left(3x^2 + \frac{1}{x}\right). \end{aligned}$$

$f(x) = x^x$

$f(x) = \exp(\ln(x^x)) = \exp(x \cdot \ln(x))$

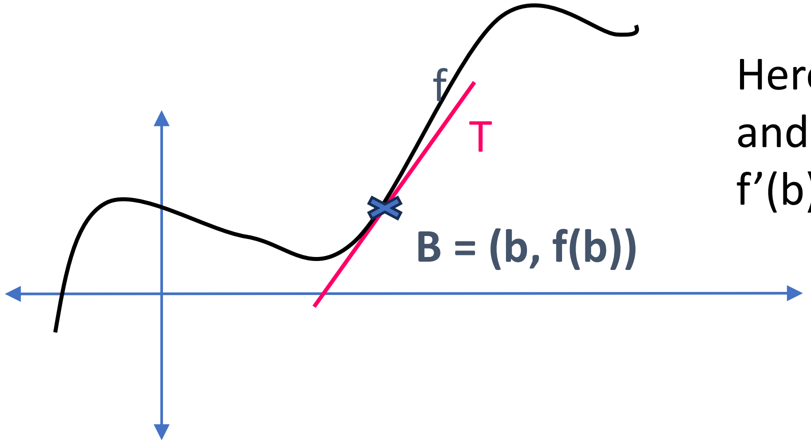
$f'(x) = \exp(x \cdot \ln(x)) \cdot (x \cdot \ln(x))' = \exp(x \cdot \ln(x)) \cdot (\ln(x) + 1) = x^x(\ln(x) + 1)$

Derivatives and extrema



tangent T of **f** at point A (a, f(a))
has equation

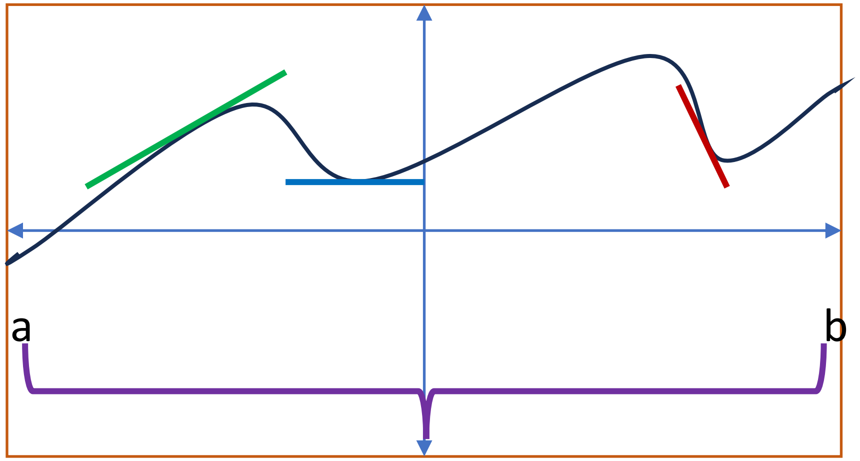
$$T(x) = f'(a)(x - a) + f(a)$$



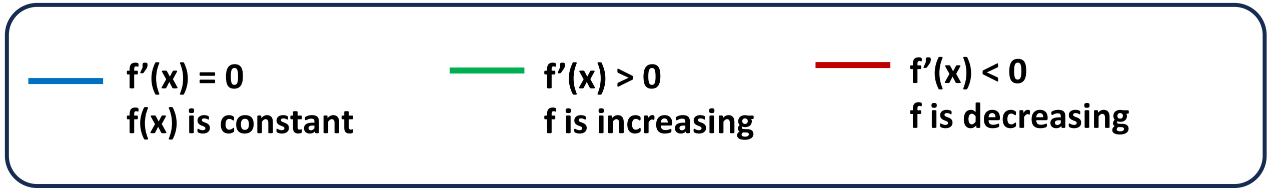
Here the slope is positive
and larger (steeper)
f'(b) > f'(a)

Derivatives and extrema

Q: Where can the maximum and minimum values (extrema) of this function be?



I = [a,b]



Let $f : I \to \mathbb{R}$ be a differentiable function. Consider an interval (a, b) contained within I, where $a < b$.

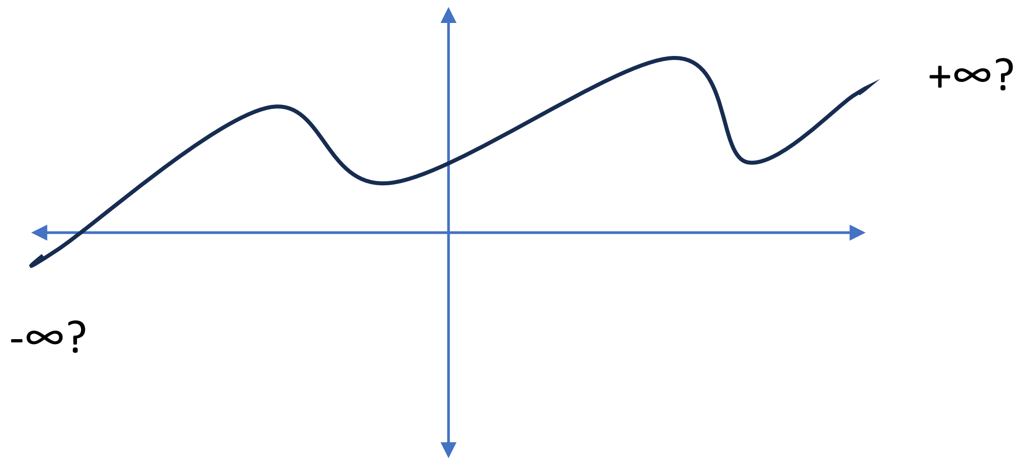
- If $f'(x)$ is (strictly) positive on (a, b), then f is (strictly) increasing on this interval.
- If $f'(x)$ is (strictly) negative on (a, b), then f is (strictly) decreasing on this interval.
- $f'(x)$ is zero on (a, b) if and only if f is constant on this interval.

A: The extrema are necessarily at

- **f(a) and/or f(b)**
- **and/or f'(x) when f'(x) = 0**

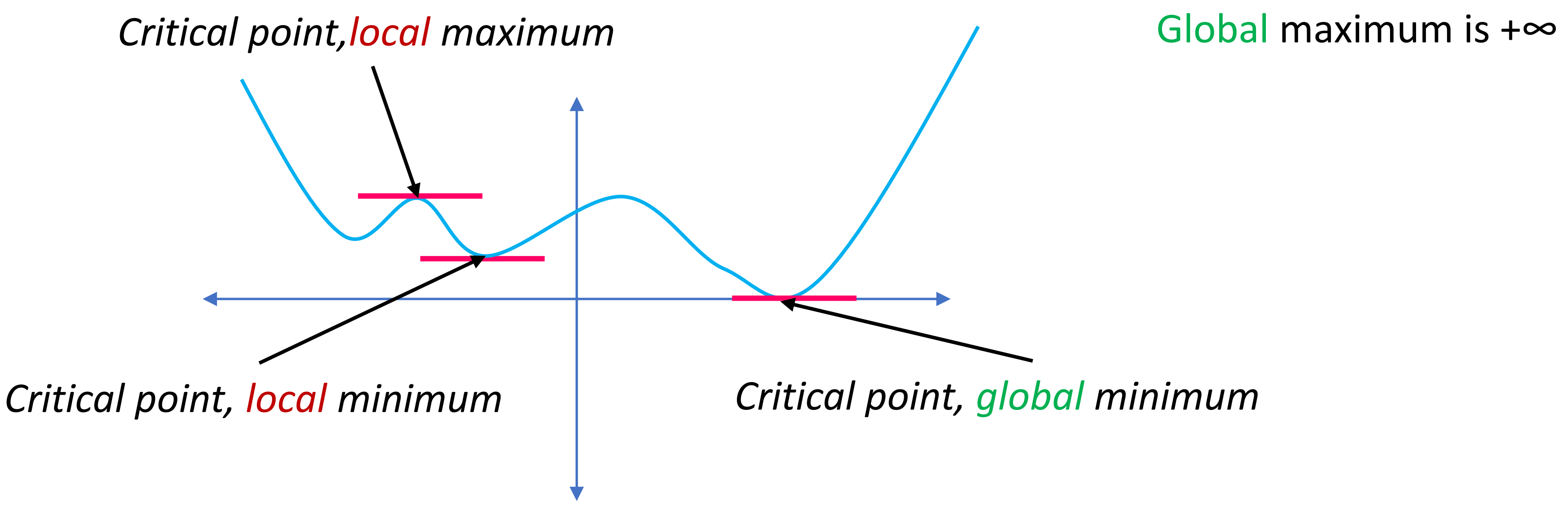
Optimization

- If we are in an open set I = (a,b), then we have no guarantee of the existence of global extrema if the function is not bounded



Optimization – no shortcuts

- critical points (**where f'(x) = 0**) are **not necessarily global extrema**



- However at local extrema, we have f'(x) = 0

Higher order derivatives

Original Function: $f(x) = \exp(x^2)$

First Derivative: $f'(x) = \exp(x^2)' = 2x\exp(x^2)$

$$
\begin{aligned}
\text{Second Derivative: } f''(x) = (2x\exp(x^2))' &= 2x\exp(x^2)' + (2x)'\exp(x^2) \\
&= 4x^2\exp(x^2) + 2\exp(x^2)
\end{aligned}
$$

Multivariable differential calculus

$$\mathbb{R} \mapsto \mathbb{R}$$
$$x \mapsto f(x)$$

From single to multivariable

$$\mathbb{R}^n \mapsto \mathbb{R}^p$$
$$x \mapsto f(x)$$

x is a **number**
f(x) is a number

x is a **vector**
f(x) is a vector

Example

$$f(x) = 2x + 3$$

Example

$$f : \mathbb{R}^2 \to \mathbb{R},\ f(x, y) = x^2 + y^4 \sin(ye^x)$$

$$g : \mathbb{R}^2 \to \mathbb{R}^3$$
$$g(x, y) = \begin{bmatrix} x^2 + y^4 \sin(ye^x) \\ y + x \\ \ln(xy) \end{bmatrix}$$

Multivariable differential calculus

Let f be a function defined on an open set U in $\mathbb{R}^n$ with values in $\mathbb{R}^p$. We denote the canonical basis of $\mathbb{R}^n$ as $\{e_1, \ldots, e_n\}$, and we fix $k \in \{1, n\}$. Given $a \in U$, we say that f has a partial derivative with respect to its k-th variable at the point a if the following quotient (where t is a real number)

$$\frac{f(a + te_k) - f(a)}{t} = \frac{f(a_1, \ldots, a_{k-1}, a_k + t, a_{k+1}, \ldots, a_n) - f(a)}{t}$$

has a limit as t approaches 0. When this limit exists, we denote it as $\frac{\partial f}{\partial k}(a)$, or simply $\frac{\partial f}{\partial x_k}(a)$ in the case where the variables are denoted as $(x_1, \ldots, x_n)$.

For example, $n = 3$ and f depends on three variables x, y, z. stating that it has a partial derivative with respect to its **second variable** at the point $(2, 1, 0)$ means that the following limit exists:

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

Multivariable differential calculus – learn by example

$$g : \mathbb{R}^2 \mapsto \mathbb{R}^3$$

$$g(x,y) = \begin{pmatrix} x^2 + y^4 \\ \sin(ye^x) \\ x + y \end{pmatrix} = \begin{pmatrix} g1(x,y) \\ g2(x,y) \\ g3(x,y) \end{pmatrix}$$

$$\frac{\partial g}{\partial x}(x,y) = \begin{pmatrix} \frac{\partial g1}{\partial x} \\ \frac{\partial g2}{\partial x} \\ \frac{\partial g3}{\partial x} \end{pmatrix} = \begin{pmatrix} 2x \\ ye^x \cos(ye^x) \\ 1 \end{pmatrix}$$

$$\frac{\partial g}{\partial y}(x,y) = \begin{pmatrix} \frac{\partial}{\partial y}(x^2 + y^4) \\ \frac{\partial}{\partial y}(\sin(ye^x)) \\ \frac{\partial}{\partial y}(x + y) \end{pmatrix} = \begin{pmatrix} 4y^3 \\ e^x \cos(ye^x) \\ 1 \end{pmatrix}$$

Multivariable differential calculus - Jacobian

Consider a function f defined on an open set U in $\mathbb{R}^n$ with values in $\mathbb{R}^p$, which has partial derivatives with respect to all its variables at a point $a \in U$. The Jacobian matrix of f at the point a, denoted as $Jf(a)$, is defined as follows:

$$Jf(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \dots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_p}{\partial x_1}(a) & \frac{\partial f_p}{\partial x_2}(a) & \dots & \frac{\partial f_p}{\partial x_n}(a) \end{pmatrix}$$

Here, for $1 \leq k \leq p$, the functions f_k are the components of the vector-valued function f.

Multivariable differential calculus

Derivatives w.r.t. all variables for g1

$$Jg((x,y)) = \begin{pmatrix} \frac{\partial}{\partial x}(x^2+y^4) & \frac{\partial}{\partial y}(x^2+y^4) \\ \frac{\partial}{\partial x}(\sin(ye^x)) & \frac{\partial}{\partial y}(\sin(ye^x)) \\ \frac{\partial}{\partial x}(x+y) & \frac{\partial}{\partial y}(x+y) \end{pmatrix} = \begin{pmatrix} 2x & 4y^3 \\ ye^x \cos(ye^x) & e^x \cos(ye^x) \\ 1 & 1 \end{pmatrix}$$

Derivatives w.r.t. y for g1, g2, g3

The Jacobian matrix of a function from $\mathbb{R}^n$ to $\mathbb{R}^p$ is an $p \times n$ matrix.

In our specific example where you are going from $\mathbb{R}^2$ to $\mathbb{R}^3$, the Jacobian matrix will be a 3×2 matrix.

gradient

The vector of all partial derivatives for function mapping to **R** is called the **gradient**

Multivariable differential calculus

Let f be a function defined on an open set U in $\mathbb{R}^n$ with values in $\mathbb{R}$, which has partial derivatives at the point $a \in \mathbb{R}^n$. The column vector $\nabla f(a)$, defined as:

$$\nabla f(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{pmatrix}$$

is called the gradient of f at a.

Hessian matrix

The second-order partial derivative of a function f that maps from $\mathbb{R}^n$ to $\mathbb{R}$ with respect to two variables x_i and x_j is denoted as $\frac{\partial^2 f}{\partial x_i \partial x_j}$ and is defined as:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)$$

The Hessian matrix of a function f is denoted as Hf and is defined as an $n \times n$ matrix where each element (i, j) is the second-order partial derivative $\frac{\partial^2 f}{\partial x_i \partial x_j}$. It is given by:

$$Hf = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Hessian matrix – example

Let's consider the function $f(x, y) = x^2 + y^2$.
To compute the Hessian matrix, we first find the first partial derivatives:

$$\frac{\partial f}{\partial x} = 2x$$

gradient

$$\frac{\partial f}{\partial y} = 2y$$

Now, let's calculate the second partial derivatives:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(2x) = 2$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}(2y) = 2$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial y}(2x) = 0$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial x}(2y) = 0$$

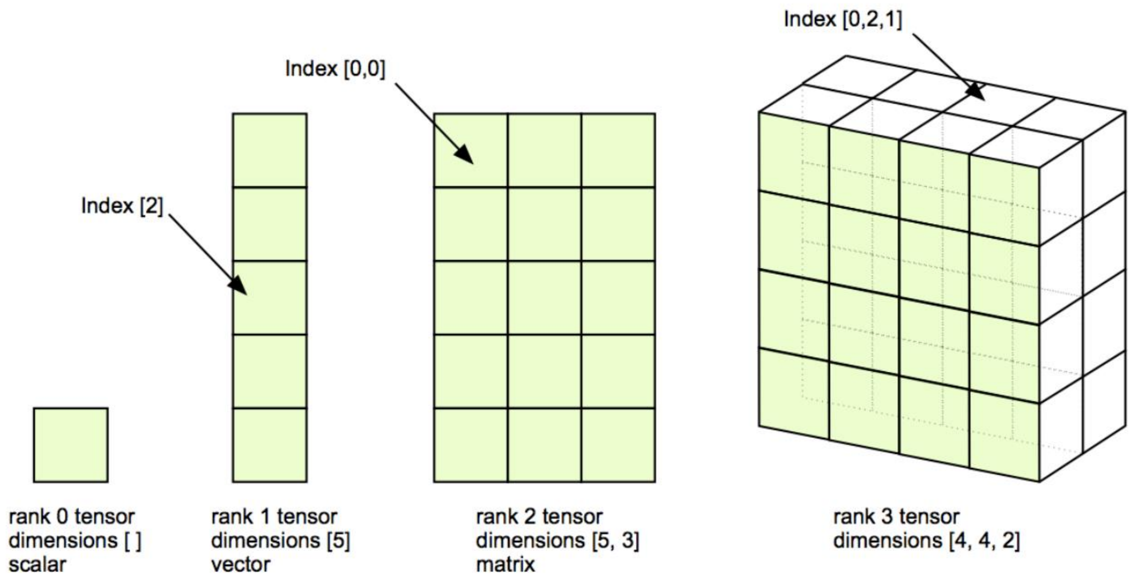
Now, we can assemble the Hessian matrix:

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Don’t get confused

Table 1: Gradient and Hessian vs. Derivative and Second Derivative

Domain	$\mathbb{R} \to \mathbb{R}$	$\mathbb{R}^p \to \mathbb{R}$	$\mathbb{R}^p \to \mathbb{R}^n$
First order	Derivative $f'(x)$ scalar	Gradient vector $\nabla f(\mathbf{x})$	Jacobian matrix $J_f(\mathbf{x})$
Second order	Second Derivative $f''(x)$ (scalar)	Hessian matrix $H_f(\mathbf{x})$	Hessian tensor $\mathbf{H}_f$



Tensors

A little bit of vector calculus

Remember the matrix-vector product $A\mathbf{x}$:

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \langle a_1, \mathbf{x} \rangle \\ \langle a_2, \mathbf{x} \rangle \\ \vdots \\ \langle a_n, \mathbf{x} \rangle \end{bmatrix}$$

What if we want to differentiate $A\mathbf{x}$ with respect to $\mathbf{x}$, which is a vector and not a scalar anymore.

A little bit of vector calculus

Vector Calculus Rules

In the following rules, $\mathbf{x}$ represents a vector. $J_f(a)$ is the Jacobian of f at x.

$$\frac{d(\mathbf{u}^T\mathbf{v})}{d\mathbf{u}} = \mathbf{v}$$

$$\frac{d(A\mathbf{x})}{d\mathbf{x}} = A$$

$$\frac{d(\mathbf{x}^T A\mathbf{x})}{d\mathbf{x}} = (A + A^T)\mathbf{x}$$

$$\frac{d(||\mathbf{x}||_2^2)}{d\mathbf{x}} = \frac{d(\mathbf{x}^T\mathbf{x})}{d\mathbf{x}} = 2\mathbf{x}$$

$$J_{f\circ g}(\mathbf{x}) = J_f(g(\mathbf{x})) \cdot J_g(\mathbf{x})$$

Equivalence with Calculus Rules

Make connections with the case where x represents a scalar

$$\frac{d(xy)}{dx} = y$$

$$\frac{d(ax)}{dx} = a$$

$$\frac{dax^2}{dx} = 2ax$$

$$\frac{dx^2}{dx} = 2x$$

$$\frac{df(g(x))}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}$$

Example Linear regression – Loss function

$$\hat{\mathbf{Y}} = \mathbf{X}\boldsymbol{\beta}$$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

N,1 N,p p,1

$$\begin{aligned} L &= \sum (y_i - \hat{y}_i)^2 \\ &= \|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2 \\ &= \|\mathbf{Y} - \mathbf{X}\beta\|_2^2 \end{aligned}$$

$$\frac{\partial L}{\partial \beta} = -2(\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{X} = -2\mathbf{Y}^T\mathbf{X} + 2\beta^T\mathbf{X}^{\mathbf{T}}\mathbf{X}$$

Transpose because we want Jacobian = gradient transposed

Optimization: finding extrema of functions

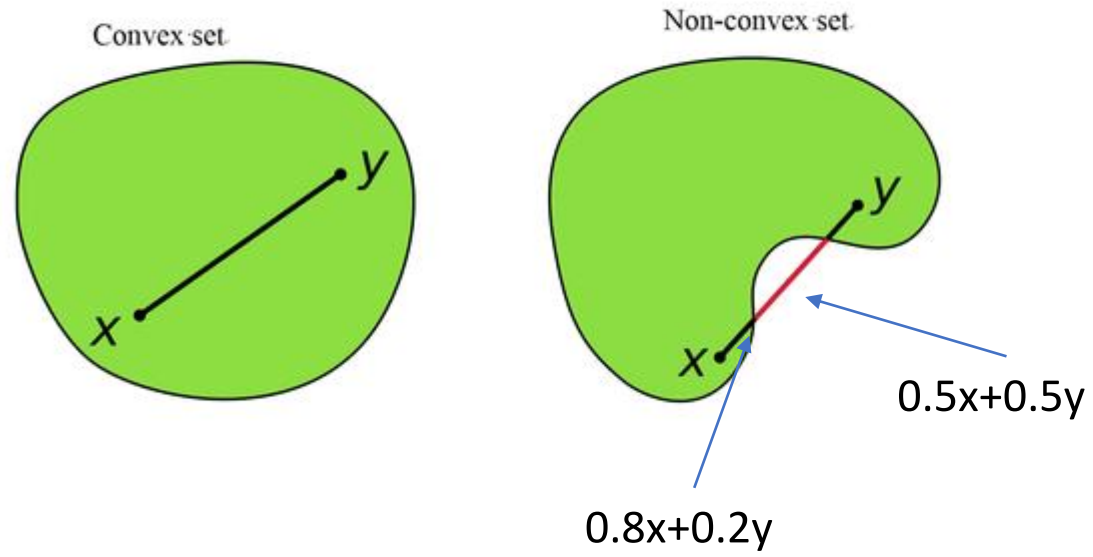
Convex Sets

Convex Set

A set S is convex if, for any two points x and y in S, the line segment connecting x and y is also contained in S.

$\forall$ x,y $\in S$, $\forall \lambda \in [0, 1]$

$$\lambda x + (1 - \lambda) y \in S$$



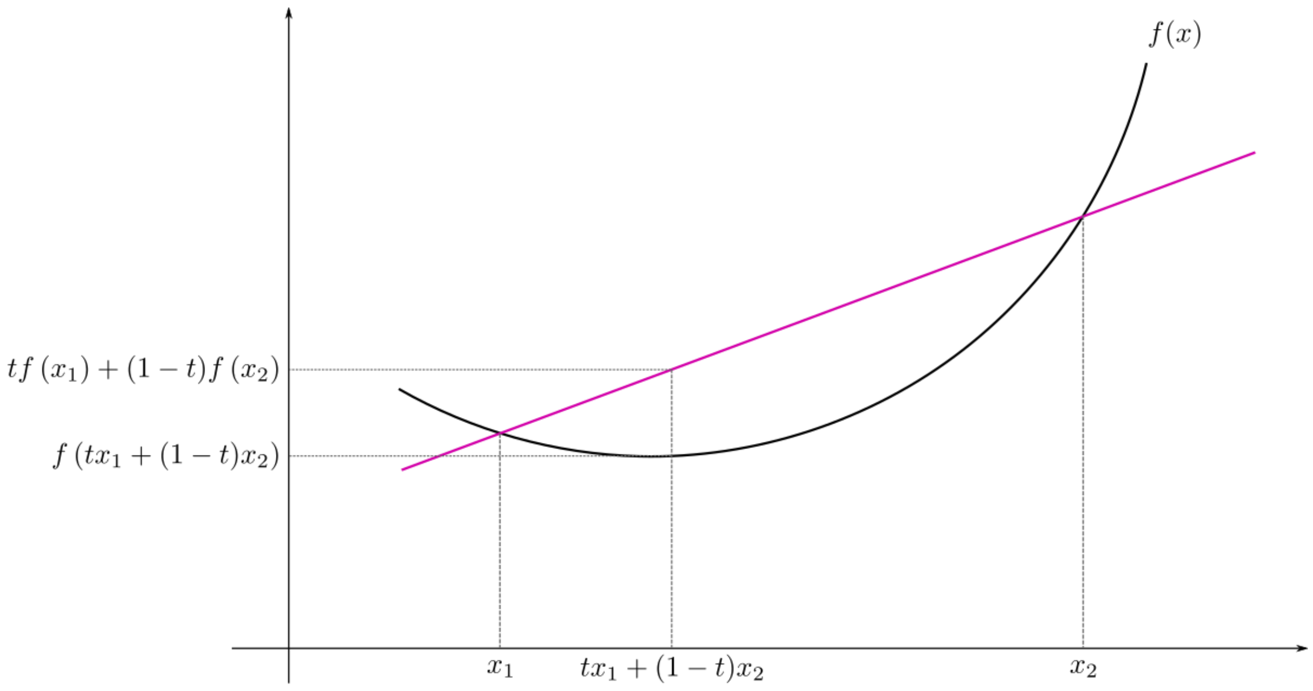
Convex functions

A function $f(x)$ is considered **convex** if, for any two points x_1 and x_2 in its domain and for any t in the interval $[0, 1]$, the following inequality holds:

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

A function $f(x)$ is considered **concave** if, for any two points x_1 and x_2 in its domain and for any t in the interval $[0, 1]$, the following inequality holds:

$$f(tx_1 + (1-t)x_2) \geq tf(x_1) + (1-t)f(x_2)$$



Convexity/concavity

Convexity = acceleration
Concavity = deceleration

- A function $f(x)$ is considered **convex** if $f''(x) \geq 0, \forall x \in \text{Dom}(f)$
- A function $f(x)$ is considered **strictly convex** if $f''(x) > 0, \forall x \in \text{Dom}(f)$
- A function $f(x)$ is considered **concave** if $f''(x) \leq 0, \forall x \in \text{Dom}(f)$
- A function $f(x)$ is considered **strictly concave** if $f''(x) < 0, \forall x \in \text{Dom}(f)$

$\exp(x)$ is **strictly convex**.

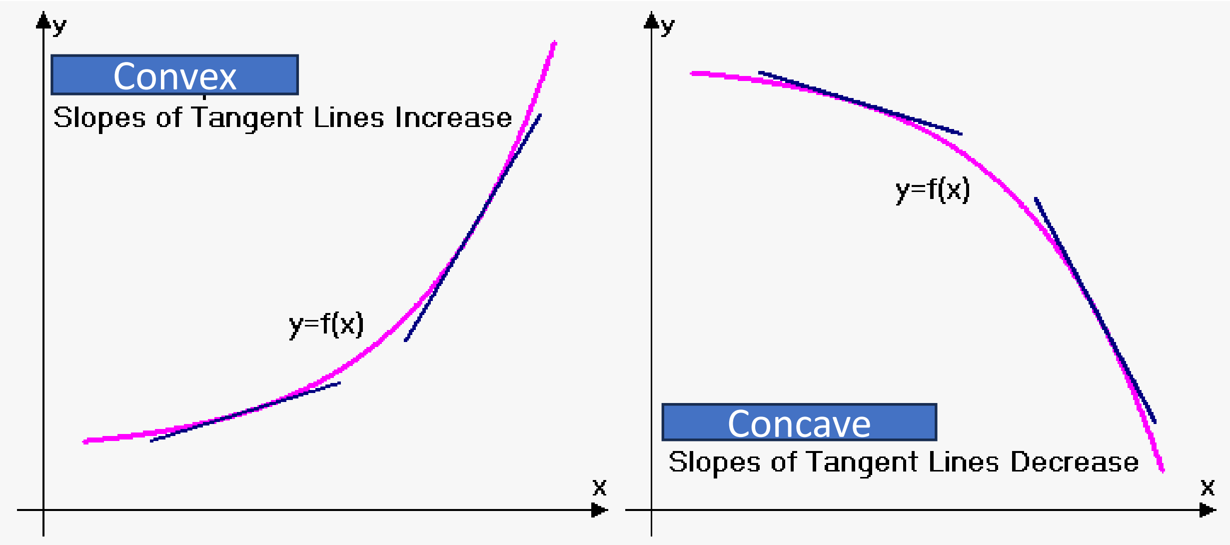
$$\exp(x)'' = \exp(x) > 0$$

x^2 is **strictly convex**.

$$(x^2)'' = 2 > 0$$

$\ln(x)$ is **strictly concave**.

$$\ln(x)'' = -\frac{1}{x^2} < 0$$



Hessian matrix and convexity

- A Hessian matrix H is positive semidefinite ($H \succeq 0$) $\iff$ all eigenvalues of $H_f \geq 0$.
- A Hessian matrix H is positive definite ($H \succ 0$) $\iff$ all eigenvalues of $H_f > 0$.
- A function f is convex $\iff$ its Hessian matrix H_f is positive semidefinite (P.S.D.)
- A function f is strictly convex $\iff$ its Hessian matrix H_f is positive definite (P.D.)

From differential calculus to optimization

Method Let $O \subset X$ be an open set. If $f : O \to \mathbb{R}$ is a differentiable function, then the local and global minimizers of f (if they exist) are among the critical points of f. Furthermore, if f is twice differentiable, then for any critical point x^* of f:

- If Hess $f(x^*)$ is positive definite, then x^* is a local minimizer of f.
- If Hess $f(x^*)$ is not positive semi-definite, then x^* is not a local minimizer of f.
- If Hess $f(x^*)$ is positive semi-definite but not positive definite, then we cannot conclude.

We do not assume any prior knowledge about the convexity of f:

1. Solve for x in $\nabla f(x^*) = 0$, where x^* is a critical point.
2. Evaluate Hess(f) at the critical point x^*.
3. Conclude based on the eigenvalues of $\text{Hess}_f(x^*)$:
   - If all eigenvalues are positive, then x^* is a local minimizer of f.
   - If any eigenvalue is negative, then x^* is not a local minimizer of f.
   - If there are zero eigenvalues (indicating semi-definiteness), further analysis is needed to make a conclusion.

Example

$$f : \mathbb{R} \to \mathbb{R}$$
$$t \mapsto t^3 + 6t^2 - 15t + 1$$

1. Find critical points

The function f is a polynomial function, and therefore it is differentiable, with its derivative f' defined as:

$$f' : \mathbb{R} \to \mathbb{R}, \quad t \mapsto 3t^2 + 12t - 15$$

The critical points of f, if they exist, are real numbers t that satisfy $f'(t) = 0$, which can be expressed as:

$$3t^2 + 12t - 15 = 0$$

The discriminant of this quadratic polynomial is $\Delta = b^2 - 4ac$, where $a = 3$, $b = 12$, and $c = -15$:

$$\Delta = 12^2 - 4 \cdot 3 \cdot (-15) = 144 + 180 = 324$$

Since $\Delta > 0$, it indicates that f has two distinct critical points, which can be found using the quadratic formula:

$$t = \frac{-b \pm \sqrt{\Delta}}{2a}$$

Thus, the correct critical points of f are:

$$t_1 = \frac{-12 + \sqrt{324}}{2 \cdot 3} = \frac{-12 + 18}{6} = \frac{6}{6} = 1$$

$$t_2 = \frac{-12 - \sqrt{324}}{2 \cdot 3} = \frac{-12 - 18}{6} = \frac{-30}{6} = -5$$

Therefore, the correct critical points of f are $t_1 = 1$ and $t_2 = -5$. .

Example

$$f : \mathbb{R} \to \mathbb{R}$$
$$t \mapsto t^3 + 6t^2 - 15t + 1$$

2. Evaluate second order condition at crit(f) = x*

The function f is a polynomial function and is twice differentiable, with its second derivative given by:

$$\forall t \in \mathbb{R}, \quad f''(t) = 6t + 12$$

f has two critical points 1 and -5.

The second derivative $f''(t)$ is positive for all $t > -2$ and negative for all $t < -2$: f is not convex.

- $f(t_1) = f''(-5) < 0$ implies that t_1 is not a local minimizer.
- $f(t_2) = f''(1) > 0$ implies that t_2 is a local minimizer.

Convexity and minimization

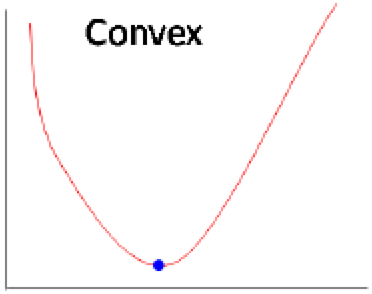
For a convex function f, the set of critical points $\text{crit}(f) := \{x \mid f'(x) = 0\}$ is equal to the set of global minimizers.

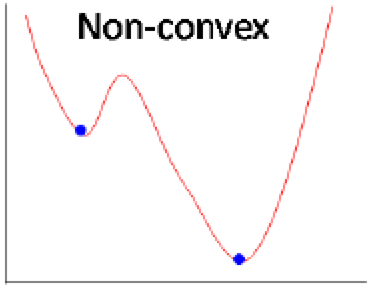
Let $f : X \to \mathbb{R}$ a convex differentiable function. Let $x^* \in X$. Then :

1. x^* is a minimizer of $f \iff \nabla f(x^*) = 0$
2. f has at most one minimizer x^* (unique if it exists)

According to this proposition, every convex function f has the interesting property of having an identity between the following three sets (which may be empty):

- the set of its global minimizers $\arg\min_x f$;
- the set of its local minimizers;
- the set crit_f of its critical points.





Example

$$f : \mathbb{R} \to \mathbb{R}$$
$$t \mapsto \sqrt{1+t^2}$$

$$f'(t) = \frac{t}{\sqrt{1+t^2}}$$
$$f''(t) = \frac{1}{\sqrt{1+t^2}(1+t^2)} > 0$$

So, f is **strictly convex**
its **unique global minimum is its critical point**. Let's find it

$$f'(t^*) = 0 \iff \frac{t^*}{\sqrt{1+t^{*2}}} = 0$$
$$\iff t^* = 0$$

The minimizer of f is $t^* = 0$ (argmin f).
The minimum value of f is $f(t^*) = 1$.

Example Linear regression – Loss function

$$\hat{\mathbf{Y}} = \mathbf{X}\boldsymbol{\beta}$$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

N,1 N,p p,1

$$L = \sum (y_i - \hat{y}_i)^2$$
$$= \|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2$$
$$= \|\mathbf{Y} - \mathbf{X}\beta\|_2^2$$

$$\frac{\partial L}{\partial \beta} = -2(\mathbf{Y} - \mathbf{X}\beta)^T\mathbf{X} = -2\mathbf{Y}^T\mathbf{X} + 2\beta^T\mathbf{X}^{\mathbf{T}}\mathbf{X}$$

Transpose because we want Jacobian = gradient transposed

Example Linear regression – Loss function

$$
\begin{aligned}
L &= \sum (y_i - \hat{y}_i)^2 \\
&= \|Y - \hat{Y}\|_2^2 \\
&= \|Y - \mathbf{X}\beta\|_2^2
\end{aligned}
$$

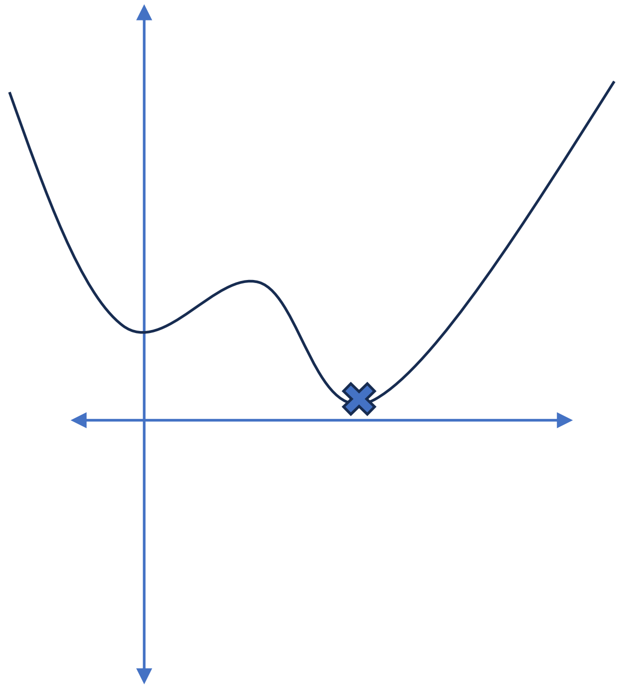
L is convex, so its critical point β^* is its global minimizer

$$
\begin{aligned}
\frac{\partial L}{\partial \beta} = 0 &\iff -2Y^T\mathbf{X} + 2\beta^T\mathbf{X^T}\mathbf{X} = 0 \\
&\iff Y^T\mathbf{X} = \beta^T\mathbf{X^T}\mathbf{X} \\
&\iff \beta^T = Y^T\mathbf{X}(\mathbf{X^T}\mathbf{X})^{-1} \\
&\iff \beta = ((\mathbf{X^T}\mathbf{X})^{-1})^T\mathbf{X^T}Y \\
&\iff \beta = (\mathbf{X}^{-1}(\mathbf{X}^T)^{-1})^T\mathbf{X^T}Y \\
&\iff \beta = (\mathbf{X}^{-1}(\mathbf{X}^T)^{-1})\mathbf{X^T}Y \\
&\iff \beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X^T}Y
\end{aligned}
$$

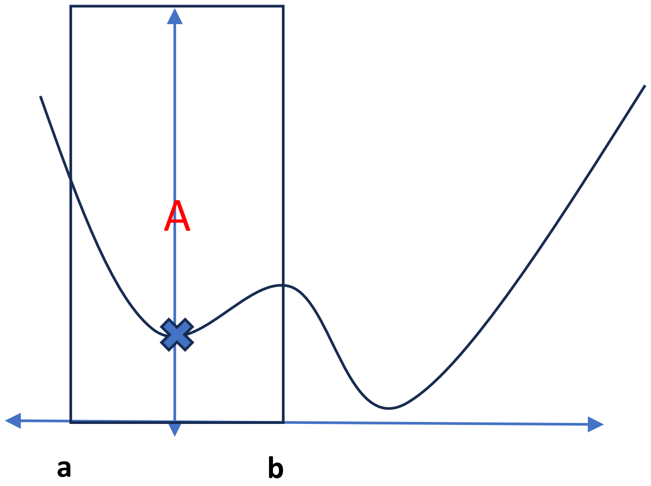
$$
\begin{aligned}
||x||_2^2 &= \sum_i x_i^2 = x^T x \\
(AB)^T &= B^T A^T \\
(A+B)^T &= A^T + B^T \\
(AB)^{-1} &= B^{-1}A^{-1} \\
(A^T)^{-1} &= (A^{-1})^T \\
(A^T)^T &= A
\end{aligned}
$$

Constrained optimization

Unconstrained:
We want x* that minimizes f(x)
x* can be anywhere in R



Constrained
We want x* that minimizes f(x)
x* is in a specific subset **A: [a;b]**



Here the constraint is $a < x < b$
it is an **inequality** constraint

Lagrangian

In optimization, the Lagrangian ($\mathcal{L}$) is a function used to formulate and solve constrained optimization problems. It is defined as follows for an objective function $f(x)$ subject to equality and inequality constraints:

For a minimization problem:

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i \cdot g_i(x) + \sum_{j=1}^{n} \mu_j \cdot h_j(x)$$

In this expression:

- x represents the vector of optimization variables.
- λ_i (Lagrange multipliers) are associated with the equality constraints $g_i(x) = 0$.
- μ_j (Lagrange multipliers) are associated with the inequality constraints $h_j(x) \leq 0$.

Lagrangian example

- **Objective Function**: We want to maximize the function $f(x, y) = 2x + 3y$.
- **Equality Constraint**: Our equality constraint is $g(x, y) = x^2 + y^2 = 4$, representing a circle with radius 2 centered at the origin.
- **Inequality Constraint**: Our inequality constraint is $h(x, y) = x - y \geq -1 \iff h(x, y) = -x + y - 1 \leq 0$.

  The Lagrangian for this problem, considering both equality and inequality constraints, is defined as:

$$\begin{aligned}\mathcal{L}(x, y, \lambda, \mu) &= f(x, y) + \lambda \cdot g(x, y) + \mu \cdot h(x, y) \\ &= 2x + 3y + \lambda(x^2 + y^2 - 4) + \mu(-x + y - 1)\end{aligned}$$

  Here, λ and μ are the Lagrange multipliers associated with the equality and inequality constraints, respectively.

KKT conditions

A point x^* satisfies the Karush-Kuhn-Tucker (KKT) conditions if there exist Lagrange multipliers $\lambda_1, \ldots, \lambda_p, \mu_1, \ldots, \mu_q$ such that:

$$\nabla \mathcal{L}(x^*) = \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{q} \mu_j \nabla h_j(x^*) = 0$$

where

- for all $i \in [1, p]$ $g_i(x^*) = 0$
- for all $j \in [1, q]$ $h_j(x^*) \leq 0$
- $\mu_j \geq 0$
- $\mu_j h_j(x^*) = 0.$

x^* is a critical point of $\mathcal{L}$ not of f

First order conditions for convex problems

Let $U \subset \mathbb{R}^n$ be an open set. Consider functions $f : U \to \mathbb{R}$ and $h_j : U \to \mathbb{R}$, for $j \in [1, q]$, which are differentiable and convex, and functions $g_i : U \to \mathbb{R}$, for $i \in [1, p]$, which are affine.

We are concerned with the following constrained optimization problem:

$$
\begin{array}{lll}
\text{Minimize} & f(x) & \\
\text{subject to the constraints} & g_i(x) = 0 \quad \text{for } i \in [1, p] & \\
 & h_j(x) \le 0 \quad \text{for } j \in [1, q] &
\end{array}
\tag{P}
$$

We say this problem is a **convex optimization problem** as the objective function is convex, the inequality constraints are convex, and the equality constraints are affine.

Idea: For a convex optimization problem, a critical point of $\mathcal{L}$ satisfying the KKT conditions is a solution, under additional conditions...

Sequences and Series

Sequences

Sequence:

A *sequence* is an ordered list of numbers denoted as $\{a_n\}$, where a_n represents the n-th term of the sequence. In general, a sequence can be defined as a function from the set of natural numbers ($\mathbb{N}$) to the set of real numbers ($\mathbb{R}$).

Arithmetic Sequence:

An *arithmetic sequence* is a sequence in which the difference between any two consecutive terms is constant. The n-th term of an arithmetic sequence can be defined as:

$$u_n = u_0 + nr$$

where u_0 is the first term, r is the common difference between consecutive terms, and n is the position of the term in the sequence.

Geometric Sequence:

A *geometric sequence* is a sequence in which the ratio of any two consecutive terms is constant. The n-th term of a geometric sequence can be defined as:

$$u_n = u_0 \cdot r^n$$

Sequences

Ex: $u_n = u_0 + 2n,\ u_0 = 3, r = 2,$ Arithmetic Sequence

$$u_0 = 3, \quad u_1 = 5, \quad u_2 = 7, \quad u_3 = 9, \quad u_4 = 11, \dots$$

Ex: $u_n = 2 \cdot 3^n\ u_0 = 2, r = 3,$ Geometric Sequence

$$u_0 = 2, \quad u_1 = 6, \quad u_2 = 18, \quad u_3 = 54, \quad u_4 = 162, \dots$$

Sequences

The sum of the first n terms of an arithmetic sequence can be calculated using the following formula:

$$S_n = u_0 + u_1 + ... + u_n = (u_0 + u_n)\frac{n+1}{2}$$

n+1 terms in the sum

where S_n is the sum of the first n terms

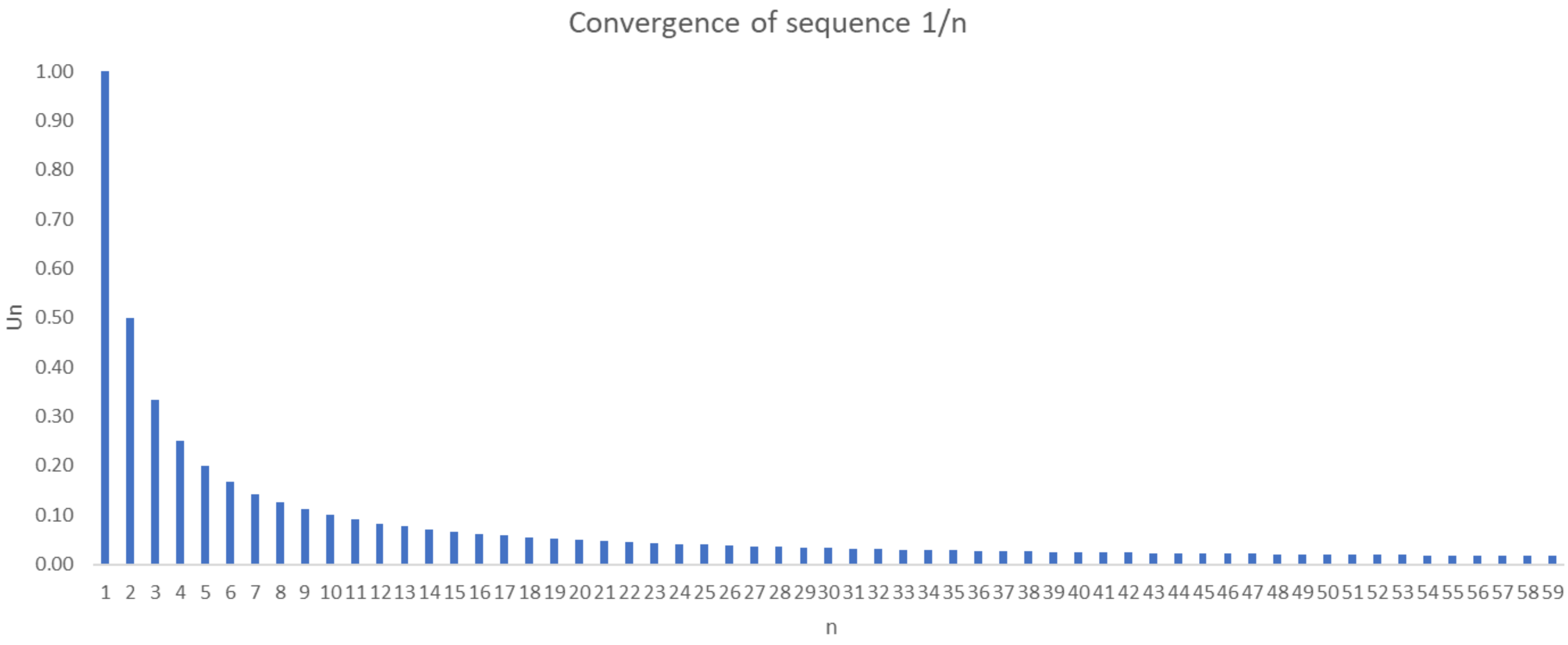
The sum of the first n terms of a geometric sequence can be calculated using the following formula:

$$S_n = u_0 \frac{1 - r^{n+1}}{1 - r}$$

Sequences

A sequence (u_n) converges if there exists $\lambda \in \mathbb{C}$ such that for all $\epsilon > 0$, there exists a rank $N \in \mathbb{N}$ from which the sequence values stay within radius $D(\lambda, \epsilon)$. Formally :

$$\exists \lambda \in \mathbb{C}, \quad \forall \epsilon > 0, \quad \exists N \in \mathbb{N}, \quad \forall n \geq N, \quad |u_n - \lambda| < \epsilon$$



Series

Given a sequence (u_n), we call the series with the general term u_n the sequence:

$$S_n = u_0 + u_1 + ... + u_n = \sum_{k=0}^{n} u_k.$$

S_n is called the n-th partial sum. We write $\sum_{k=0}^{n} u_k$ or simply $\sum u_k$ to refer to the sequence whose n-th term is S_n.

Be careful!! S_n is a sequence, it is a sequence of sums of u_n, which is also a sequence

For instance if u_n has 3 terms

$$u_n = (u_0, u_1, u_2) = (1, 4, 8)$$

$$S_n = (S_0, S_1, S_2) = (u_0, u_0 + u_1, u_0 + u_1 + u_2) = (1, 5, 13)$$

Summation operator

Properties of the Summation Operator:

1. Linearity:

$$\sum_{k=m}^{n} (c \cdot a_k) = c \cdot \sum_{k=m}^{n} a_k$$

for any constant c.

2. Splitting:

$$\sum_{k=m}^{n} (a_k + b_k) = \sum_{k=m}^{n} a_k + \sum_{k=m}^{n} b_k$$

3. Changing the Index:

$$\sum_{k=m}^{n} a_k = \sum_{j=m}^{n} a_j$$

This property allows you to use a different index variable.

4. Constant Term:

$$\sum_{k=m}^{n} c = (n - m + 1) \cdot c$$

when all terms are constant.

5. Telescoping Series:

$$\sum_{k=m}^{n} (a_k - a_{k+1}) = (a_m - a_{n+1})$$

This property simplifies some series by canceling out adjacent terms.

Convergence of Series

Let (u_n) be a sequence of complex numbers. We say that $\sum_{k=0}^{\infty} u_k$ is convergent if the sequence (S_n) is convergent. If $\sum_{k=0}^{\infty} u_k$ does not converge, it is said to be divergent. If $\sum_{k=0}^{\infty} u_k$ converges, we write:

$$\sum_{k=0}^{\infty} u_k = \lim_{n \to \infty} S_n.$$

Please note that we can ONLY write the symbol $\sum_{k=0}^{\infty} u_k$ if we have already proven that $\sum u_k$ converges!!!

Convergence of Series

Let's show that the series

$$\sum_{k=0}^{\infty} \frac{1}{(k+1)(k+2)}$$

converges.

For any positive integer n, we have:

$$\sum_{k=0}^{n} \frac{1}{(k+1)(k+2)} = \sum_{k=0}^{n} \left(\frac{1}{k+1} - \frac{1}{k+2}\right) = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \ldots + \left(\frac{1}{n+1} - \frac{1}{n+2}\right) = 1 - \frac{1}{n+2}.$$

Hence, the series converges with a sum of

$$\sum_{k=0}^{\infty} \frac{1}{(k+1)(k+2)} = 1.$$

Convergence of Series

Proposition [Convergence of the Geometric Series]
Let $z \in \mathbb{C}$. Then, the series

$$\sum_{k=0}^{\infty} z^k$$

is convergent if and only if $|z| < 1$, and in that case:

$$\forall z \in \mathbb{C}, \quad |z| < 1, \quad \sum_{k=0}^{\infty} z^k = \frac{1}{1-z}.$$

Proof:
Assume that $\sum_{k=0}^{\infty} z^k$ is convergent. This implies that z^n approaches zero as n goes to infinity, and therefore, $|z|^n$ also approaches zero. Consequently, $|z| < 1$.
Conversely the sum of a geometric sequence is given by:

$$\sum_{k=0}^{n} z^k = \frac{1-z^{n+1}}{1-z}.$$

Since $|z| < 1$, we have $\lim_{n\to\infty} z^n = 0$. Thus, we obtain:

$$\forall z \in \mathbb{C}, \quad \sum_{k=0}^{\infty} z^k = \frac{1}{1-z}.$$

Power series

We call an power series any series of functions $\sum_{n=0}^{\infty} f_n$ where $f_n : z \to a_n z^n$ for $z \in \mathbb{C}$ and $a_n \in \mathbb{C}$ for $n \in \mathbb{N}$. The a_n are called the coefficients of the power series. For convenience, we write $\sum_{n=0}^{\infty} a_n z^n$ to represent such a series.

We can use **power series expansion** to express usual functions, for instance

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$
$$= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + ...$$

The factorial of a non-negative integer n, denoted as $n!$, is defined:

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$$

$0! = 1$.

For example, $5!$ is calculated as:

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

O-notations

Definition: Let x_0 be a point in $\mathbb{R}$. A neighborhood of x_0 is an open interval containing x_0. These are often taken in the form $(x_0 - \delta, x_0 + \delta)$ where $\delta > 0$.

Definitions: Let x_0 be a point in $\mathbb{R}$. Suppose f and g are two functions defined in a neighborhood of x_0, such that the function g only equals zero at the point x_0. We say that:

- f is little-o of g in the neighborhood of x_0, denoted as $f = o_{x_0}(g)$, if

f grows slower than g around x0

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0.$$

- f is equivalent to g in the neighborhood of x_0, denoted as $f \sim_{x_0} g$, if

f grows at the same rate as g around x0

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 1.$$

O-notations - example

Let $f(x) = (x-3)^2$, $g(x) = (x-3)$, and $h(x) = (x-3)^2 \exp(x-3)$.

1. f is a little-o of g in the neighborhood of $x_0 = 3$, i.e., $f = o_3(g)$. This is because:

$$\frac{f(x)}{g(x)} = \frac{(x-3)^2}{(x-3)} = (x-3),$$

and thus,

$$\lim_{x \to 3} \frac{f(x)}{g(x)} = 0.$$

2. f is equivalent to h in the neighborhood of $x_0 = 3$, i.e., $f \sim_3 h$. This is because:

$$\frac{f(x)}{h(x)} = \frac{(x-3)^2}{(x-3)^2 \exp(x-3)} = \frac{1}{\exp(x-3)},$$

and thus,

$$\lim_{x \to 3} \frac{f(x)}{h(x)} = 1.$$

Taylor Expansion

Definition: Let I be an interval in $\mathbb{R}$, and x_0 be a point or an endpoint of I. We say that a function $f : I \to \mathbb{R}$ has a Taylor expansion of order n at x_0 if there exist coefficients $a_0, a_1, \dots, a_n \in \mathbb{R}$ such that, as h tends to zero,

$$f(x_0 + h) = a_0 + a_1 h + a_2 h^2 + \ldots + a_n h^n + o_0(h^n).$$

The polynomial function $h \mapsto \sum_{i=0}^{n} a_i h^i$ of degree at most n is called the principal part of the Taylor expansion of f at x_0, and the term $o_0(h^n)$ represents the remainder of this expansion.

Theorem: Let $f : I \to \mathbb{R}$ be a smooth function and x_0 a point in the interval I. Then, for any integer n, f has a Taylor expansion of order n at x_0. This Taylor expansion is given by

$$f(x_0+h) = f(x_0)+f'(x_0)h+\frac{f''(x_0)}{2!}h^2+\ldots+\frac{f^{(n)}(x_0)}{n!}h^n+o_0(h^n) = \sum_{i=0}^{n}\frac{f^{(i)}(x_0)}{i!}h^i+o_0(h^n).$$

Maclaurin series

Taylor expansion is given by

$$f(x_0+h) = f(x_0)+f'(x_0)h+\frac{f''(x_0)}{2!}h^2+\ldots+\frac{f^{(n)}(x_0)}{n!}h^n+o_0(h^n) = \sum_{i=0}^{n}\frac{f^{(i)}(x_0)}{i!}h^i+o_0(h^n).$$

Take $x_0 = 0, h = x$ and we can approximate $f(x)$ when x is around 0. This is called the MacLaurin Series

$$f(x) = f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\ldots+\frac{f^{(n)}(0)}{n!}x^n+o_0(x^n) = \sum_{i=0}^{n}\frac{f^{(i)}(0)}{i!}x^i+o_0(x^n).$$

Animation

Maclaurin series

- For e^x:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + o_0(x^n)$$

- For $\sin(x)$:

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots + (-1)^n \frac{(2n+1)!}{(2n+1)!} x^{2n+1} + o_0(x^{2n+1})$$

- For $\cos(x)$:

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots + (-1)^n \frac{(2n)!}{(2n)!} x^{2n} + o_0(x^{2n})$$

- For $\frac{1}{1-x}$:

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots + x^n + o_0(x^n)$$

- For $\frac{1}{1+x}$:

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \ldots + (-1)^n x^n + o_0(x^n)$$

- For $\ln(1+x)$:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots + (-1)^{n+1} \frac{x^n}{n} + o_0(x^n)$$

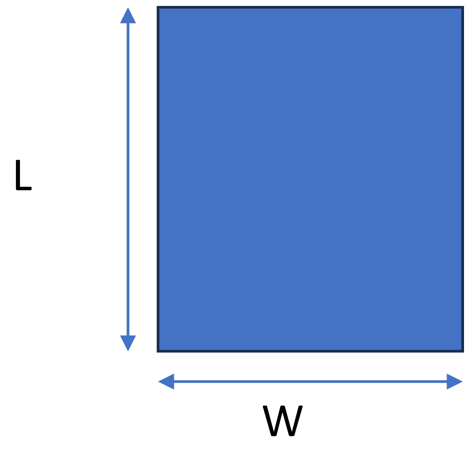
- For $(1+x)^\alpha$:

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2} x^2 + \ldots + \frac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!} x^n + o_0(x^n)$$

[Animation]

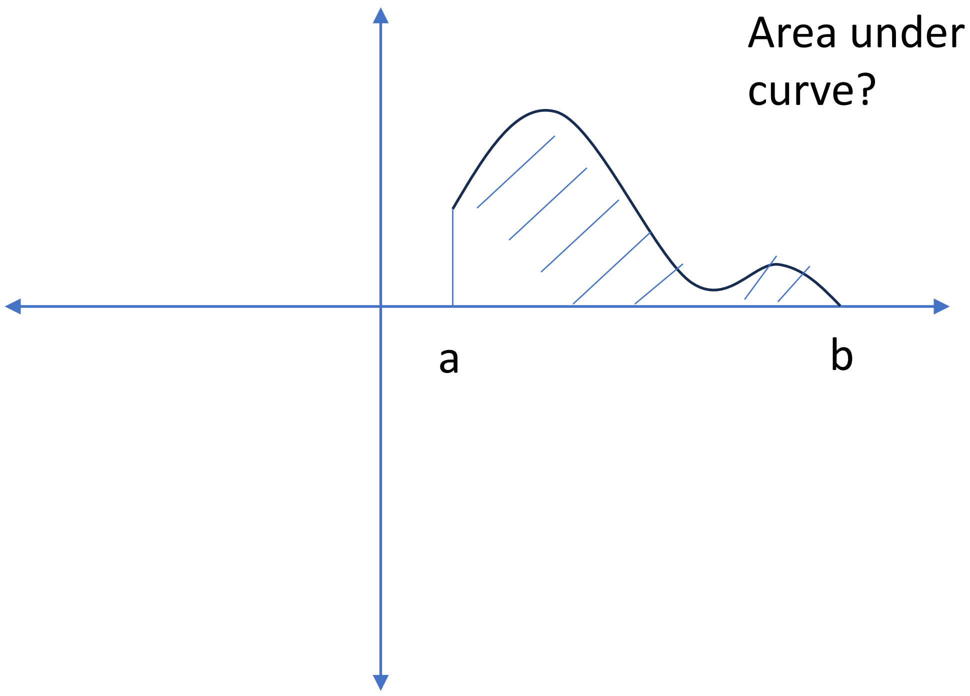
(Riemann) Integration

Integral Calculus



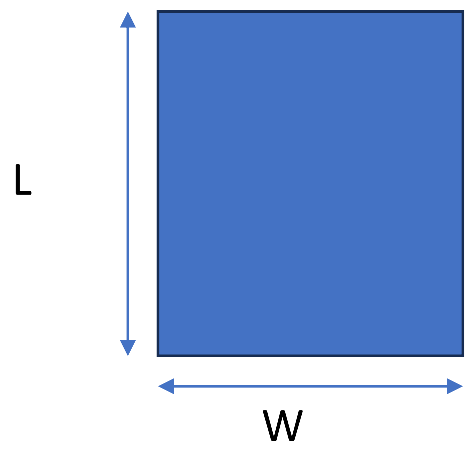
Area of this rectangle?

Area = L x W



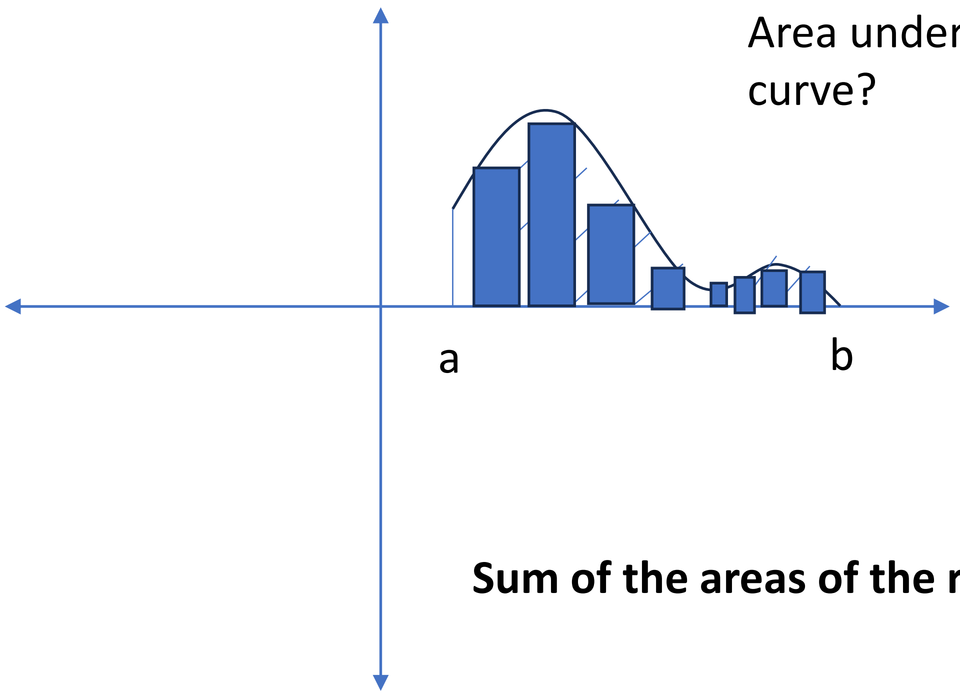
Area under this curve?

Integral Calculus



Area of this rectangle?

L x W



Area under this curve?

Sum of the areas of the rectangle?

Integral Calculus

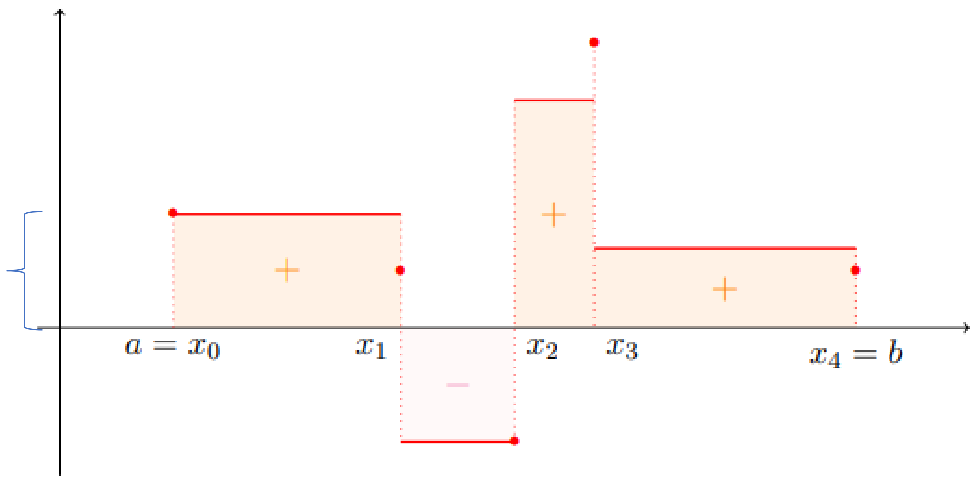
Definition: The integral of a step function is defined as the difference between, on the one hand, the sum of the areas of the rectangles formed by the step function that are located above the x-axis, and on the other hand, the sum of the areas of the rectangles located below the x-axis. In other words, if f is a step function associated with the subdivision $\sigma = \{x_0 < x_1 < \ldots < x_n\}$ of $[a, b]$, it is given by

$$\int_a^b f = \sum_{i=1}^{n} (x_i - x_{i-1}) f(m_i),$$

where $m_i = \frac{x_{i-1}+x_i}{2}$ for $i = 1, \ldots, n$.

So, it represents the area under the curve if the step function takes only positive values. Otherwise, it is an "algebraic" area: we count positively the area above the x-axis and negatively the area below it.

f(m1)



Interruption: infimum and supremum

What is the minimum value of interval A = (-1;1)?

Is it -1 ? *NO*
Is it -0.999, -0.9999, -0.999999?

For open sets, **we extend the idea of the minimum and maximum elements to inf. and sup**.

Inf(A) = -1
Sup(A) = 1

Interruption: infimum and supremum

Sup/Inf(A) is 'sticky' to A

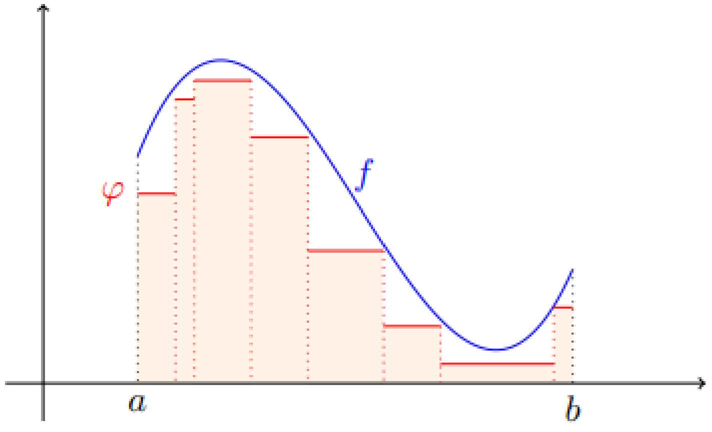
- **Supremum** ($\sup A$)**:** Every non-empty and bounded subset A of $\mathbb{R}$ has a least upper bound, denoted as $\sup A$. This is the smallest of the upper bounds, meaning it is the unique real number satisfying the following two properties:
  - For all $a \in A$, $a \leq \sup A$.
  - For every $\epsilon > 0$, there exists $a \in A$ such that $a > \sup A - \epsilon$.
- **Infimum** ($\inf A$)**:** Every non-empty and bounded subset A of $\mathbb{R}$ has a greatest lower bound, denoted as $\inf A$. This is the largest of the lower bounds, meaning it is the unique real number satisfying the following two properties:
  - For all $a \in A$, $\inf A \leq a$.
  - For every $\epsilon > 0$, there exists $a \in A$ such that $\inf A + \epsilon > a$.

Riemann Integration

We can consider step functions ϕ whose graphs are below that of f: $\phi \le f$. Each of these functions ϕ has an integral, defined as an algebraic area, as described in the previous section. One way to conceive the integral of f is that it should be the largest area obtained in this manner. More precisely, we define the lower integral of f using an upper bound:

$$I^{-}_{a,b}(f) = \sup \left\{ \int_a^b \phi \mid \phi \in E([a,b]), \phi \le f \right\}.$$

We refer to it as the lower integral because we approximate the graph of f from below, using functions $\phi \le f$.

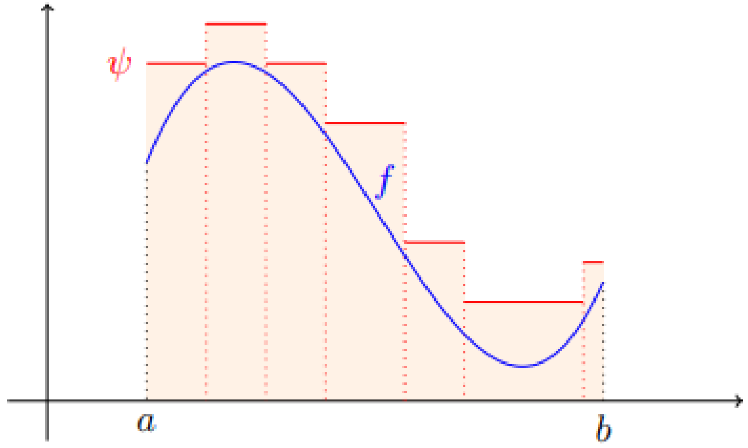


Riemann Integration

We can consider step functions ψ whose graphs are above that of f: $\psi \leq f$. Each of these functions ψ has an integral, defined as an algebraic area, as described in the previous section. One way to conceive the integral of f is that it should be the smallest area obtained in this manner. More precisely, we define the upper integral of f using an lower bound:

$$I^+_{a,b}(f) = \inf \left\{ \int_a^b \psi \mid \psi \in E([a,b]), \psi \geq f \right\}.$$

We refer to it as the upper integral because we approximate the graph of f from above, using functions $\psi \geq f$.



Riemann Integration

Let f be a bounded function on $[a, b]$. We say that f is integrable over $[a, b]$ when $I^+_{a,b}(f) = I^-_{a,b}(f)$. In this case, we denote the common value of $I^+_{a,b}(f)$ and $I^-_{a,b}(f)$ as $\int_a^b f$.

Fundamental theorem of calculus

First Fundamental Theorem of Calculus:

Let $f(x)$ be a continuous function on a closed interval $[a, b]$. If $F(x)$ is any antiderivative of $f(x)$ on $[a, b]$, then:

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

In simpler terms, this theorem states that if you can find an antiderivative $F(x)$ of a continuous function $f(x)$, then you can calculate the definite integral of $f(x)$ over the interval $[a, b]$ by evaluating $F(x)$ at the upper and lower limits of integration and subtracting the results.

Antiderivative F(x) means F'(x) = f(x)

For $f(x) = x$, the antiderivative $F(x)$ is $\frac{x^2}{2}$

Integral Calculus

Functions of x	**Antiderivatives**
Constant	$\int k\,dx = kx + C$
Power Rule	$\int x^n\,dx = \frac{1}{n+1}x^{n+1} + C$, where $n \neq -1$
Exponential Function	$\int e^x\,dx = e^x + C$
Natural Logarithm	$\int \frac{1}{x}\,dx = \ln\lvert x\rvert + C$, where $x \neq 0$
Trigonometric Functions	$\int \sin(x)\,dx = -\cos(x) + C$ $\int \cos(x)\,dx = \sin(x) + C$ $\int \frac{1}{1+x^2}\,dx = \arctan(x) + C$

Functions of u	**Antiderivatives**
Power Rule	$\int nu'u^n\,du = \frac{1}{n+1}u^{n+1} + C$, where $n \neq -1$
Exponential Function	$\int u'e^u\,du = e^u + C$
Natural Logarithm	$\int \frac{u'}{u}\,du = \ln\lvert u\rvert + C$, where $u \neq 0$

Integration = sum in a continuous setting

- **Linearity:** The integral operator is linear, meaning that for constants c_1 and c_2 and functions $f(x)$ and $g(x)$, we have:

$$\int [c_1 f(x) + c_2 g(x)]\, dx = c_1 \int f(x)\, dx + c_2 \int g(x)\, dx$$

- **Additivity:** For any three numbers a, b, and c within the interval $[a, b]$, we have:

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

- **Symmetry:** If $f(x)$ is an even function ($f(-x) = f(x)$), then for any interval symmetric about the origin ($[-a, a]$), the integral simplifies to:

$$\int_{-a}^a f(x)\, dx = 2 \int_0^a f(x)\, dx$$

Integration – example

Example: Find the integral of the function $f(x) = 2xe^{x^2}$ on the closed interval $[0, 1]$.

We want to calculate:

$$\int_0^1 2xe^{x^2}\, dx$$

To find this integral, we can apply the First Fundamental Theorem of Calculus. First, we need to find the antiderivative of $2xe^{x^2}$.

The antiderivative of $2xe^{x^2}$ is:

$$\int 2xe^{x^2}\, dx = e^{x^2} + C$$

Now, we can apply the Fundamental Theorem:

$$\int_0^1 2xe^{x^2}\, dx = \left[e^{x^2}\right]_0^1 = e^{1^2} - e^{0^2}$$

You can evaluate this numerically to find the value of the integral over the closed interval $[0, 1]$.

Double integrals

$$\int_0^1 \int_0^2 (x + 2y)\, dy\, dx$$
$$= \int_0^1 \{ \int_0^2 (x + 2y)\, dy \}\, dx$$
$$= \int_0^1 \left[xy + y^2 \right]_0^2 dx \quad \text{(Integrate with respect to } y\text{)}$$
$$= \int_0^1 (2x + 4)\, dx \quad \text{(Evaluate the limits)}$$
$$= \left[x^2 + 4x \right]_0^1 \quad \text{(Integrate with respect to } x\text{)}$$
$$= (1^2 + 4 \cdot 1) - (0^2 + 4 \cdot 0)$$
$$= 1 + 4$$
$$= 5$$

Integration by Parts

Ideally,
f has a simple integral,
g a simple derivative

So that **fg'** has a simpler integral than **f'g**

$$\int_a^b f'(x)g(x)\,dx = [f(x)g(x)]_a^b - \int_a^b f(x)g'(x)\,dx$$

Let $T > 0$ be a real number. Let's compute

$$\int_0^T te^{-t}\,dt.$$

To do this, we set $g(t) = t$ (differentiating will decrease the degree) and $f(t) = e^{-t}$. Then, we have $f'(t) = -e^{-t}$ and $g'(t) = 1$. We obtain

$$\int_0^T te^{-t}\,dt = \left[te^{-t}\right]_0^T - \int_0^T (-e^{-t})\,dt.$$

Computing $\left[te^{-t}\right]_0^T = Te^{-T}$, we are left with

$$\int_0^T (-e^{-t})\,dt = \left[e^{-t}\right]_0^T = e^{-T} - 1.$$

In conclusion, we have

$$\int_0^T te^{-t}\,dt = 1 - (T+1)e^{-T}.$$

Antiderivative of ln(x)

$$\int \ln(x)\,dx = \int \ln(x)\cdot 1\ dx$$

We pose $f'(x) = 1,\ g(x) = \ln(x)$. Then $f(x) = x,\ g'(x) = \frac{1}{x}$ using IBP:

$$\begin{aligned}\int \ln(x)\,dx &= [x\ln(x)] - \int \frac{1}{x}\cdot xdx \\ &= x\ln(x) - x + C\end{aligned}$$

Change of variable (u-sub)

Change of Variables: Under certain conditions, you can perform a change of variables to simplify an integral. For example, if g and f are differentiable functions with continuous derivatives, then:

$$\int_a^b f(g(x))g'(x)\,dx = \int_{g(a)}^{g(b)} f(u)\,du$$

Consider the integral:

$$\int_0^2 x\cos(x^2+1)\,dx.$$

Make the substitution $u = x^2+1$ to obtain $du = 2x\,dx$, meaning $dx = \frac{1}{2x}\,du$. Therefore,

$$\int_0^2 x\cos(x^2+1)\,dx = \int_1^5 x\cos(u)\frac{1}{2x}du = \frac{1}{2}\int_1^5 \cos(u)\,du = \frac{1}{2}(\sin(5)-\sin(1)).$$