1. ALGEBRA

1.1. Arithmetic Operations

$a(b+c) = ab + ac$ $\quad$ $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{db}$

$\frac{a+c}{b} = \frac{a}{b} + \frac{c}{b}$ $\quad$ $\frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a}{b} \times \frac{d}{c} = \frac{ad}{bc}$

1.2. Exponents and Radicals

$x^m x^n = x^{m+n}$ $\quad$ $\frac{x^m}{x^n} = x^{m-n}$

$(x^m)^n = x^{mn}$ $\quad$ $x^{-n} = \frac{1}{x^n}$

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

$x^{\frac{1}{n}} = \sqrt[n]{x}$ $\quad$ $x^{\frac{m}{n}} = \sqrt[n]{x^m} = (\sqrt[n]{x})^m$

$\sqrt[n]{xy} = \sqrt[n]{x}\,\sqrt[n]{y}$ $\quad$ $\sqrt[n]{\frac{x}{y}} = \frac{\sqrt[n]{x}}{\sqrt[n]{y}}$

1.3. Factoring Special Polynomials

$x^2 - y^2 = (x+y)(x-y)$

$x^3 + y^3 = (x+y)(x^2 - xy + y^2)$

$x^3 - y^3 = (x-y)(x^2 + xy + y^2)$

1.4. Binomial Theorem

$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$

$(x+y)^2 = x^2 + 2xy + y^2$ $\quad$ $(x-y)^2 = x^2 - 2xy + y^2$

$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$ $\quad$ $(x-y)^3 = x^3 - 3x^2y + 3xy^2 - y^3$

1.5. Quadratic Formula

If $ax^2 + bx + c = 0$, then $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$

1.6. Inequalities and Absolute Value

If $a < b$ and $b < c$, then $a < c$

If $a < b$, then $a + c < b + c$

If $a < b$ and $c > 0$, then $ca < cb$

If $a < b$ and $c < 0$, then $ca > cb$

If $a > 0$, then

$|x| = a$ means $x = a$ or $x = -a$

$|x| < a$ means $-a < x < a$

$|x| > a$ means $x > a$ or $x < -a$

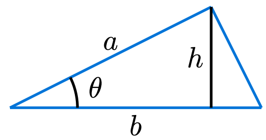
2. GEOMENTRY

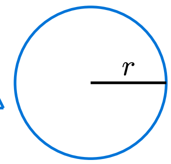
2.1. Geometric Formulas

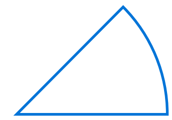
Formulas for:

- Area A
- Circumference C
- Volume V

Triangle	Circle	Sector of Circle
$A = \frac{1}{2}bh$	$A = \pi r^2$	$A = \frac{1}{2}r^2\theta$
$A = \frac{1}{2}ab\sin(\theta)$	$C = 2\pi r$	$s = r\theta$ (θ in radians)







2.2. Trigonometric Functions

Function	Definition	Reciprocal
$\sin(x)$	$\frac{\text{opposite}}{\text{hypothenuse}}$	$\frac{1}{\sin(x)} = \csc(x)$
$\cos(x)$	$\frac{\text{adjacent}}{\text{hypothenuse}}$	$\frac{1}{\cos(x)} = \sec(x)$
$\tan(x)$	$\frac{\sin(x)}{\cos(x)} = \frac{\text{opposite}}{\text{adjacent}}$	$\frac{1}{\tan(x)} = \cot(x)$
$\csc(x)$	$\frac{\sin(x)}{\cos(x)} = \frac{\text{hypothenuse}}{\text{opposite}}$	$\frac{1}{\csc(x)} = \sin(x)$
$\sec(x)$	$\frac{\text{hypothenuse}}{\text{adjacent}}$	$\frac{1}{\sec(x)} = \cos(x)$
$\cot(x)$	$\frac{\cos(x)}{\sin(x)} = \frac{\text{adjacent}}{\text{opposite}}$	$\frac{1}{\cot(x)} = \tan(x)$

Contents

Linear Algebra

Vector	$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$	
Scalar Vector Multiplication	$c \in \mathbb{R} \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ $c\vec{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix}$	
Dot Product	$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ $\begin{aligned} \vec{x} \cdot \vec{y} &= \sum_{i=1}^{n} x_i y_i \\ &= x_1 y_1 + x_2 y_2 + \ldots + x_n y_n \\ &= \vec{x}^T \vec{y} \end{aligned}$	
Cross Product ($\mathbb{R}^3$)	$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ $\vec{c} = \vec{a} \times \vec{b}$ $\vec{c} = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix}$	Returns a vector orthogonal to the two vectors
Vector Space	Closure under addition $\vec{u}, \vec{v} \in V \quad \Rightarrow \quad \vec{u} + \vec{v} \in V$ Closure under scalar multiplication $\vec{v} \in V \wedge c \in \mathbb{R} \quad \Rightarrow \quad c\vec{v} \in V$ Commutativity of addition $\vec{u} + \vec{v} = \vec{v} + \vec{u}$ Associativity of addition $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$ Additive identity $\exists \mathbf{0} \in V \mid \vec{v} + \mathbf{0} = \vec{v}$ Additive inverse $\forall \vec{v} \in V \exists -\vec{v} \in V \mid \vec{v} + (-\vec{v}) = 0$ Scalar multiplication (compatibility) $a(b\vec{v}) = (ab)\vec{v}$ Distributivity over vector addition $a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v}$	

	Distributivity over scalar addition $(a+b)\vec{v} = a\vec{v} + b\vec{v}$ Multiplicative identity $1\vec{v} = \vec{v}$	
Subspace	Non-emptiness $\mathbf{0} \in V$ Closure under addition If $\vec{u}, \vec{v} \in V$, then $\vec{u} + \vec{v} \in V$ Closure under scalar multiplication If $\vec{v} \in V, c \in \mathbb{R}$, then $c\vec{v} \in V$	A subspace is a subset of a vector space that is itself a vector space, satisfying the same axioms as the original. If V is a vector space in $\mathbb{R}^n$, then the subspace U is always contained in $\mathbb{R}^n$, meaning $U \subseteq \mathbb{R}^n$
Vector Addition	$\vec{u} = (u_1, u_2, ..., u_n)$ $\vec{v} = (v_1, v_2, ..., v_n)$ $\vec{u} + \vec{v} = (u_1 + v_1, u_2 + v_2, ..., u_n + v_n)$	
Dot Product	$\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} u_i v_i$	
Orthogonality	$\vec{u} \cdot \vec{v} = \mathbf{0}$	Angle between the two vectors is 90°
Angle between vectors	$\Theta = \arccos\left(\frac{\vec{u} \cdot \vec{v}}{\Vert\vec{u}\Vert_2 \cdot \Vert\vec{v}\Vert_2}\right)$	
L_1 Norm (Manhattan)	$\Vert\vec{u}\Vert_1 = \sum_{i=1}^{n} \vert u_i \vert$	
L_2 Norm (Euclidean)	$\Vert\vec{u}\Vert_2 = \sqrt{\sum_{i=1}^{n} u_i^2}$	
L_1 Distance (Manhattan)	$d(\vec{u}, \vec{v}) = \sum_{i=1}^{n} \vert u_i - v_i \vert$	
L_2 Distance (Euclidean)	$d(\vec{u}, \vec{v}) = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}$	
Projection	$\text{proj}_w(\vec{v}) = \frac{\vec{v} \cdot \vec{w}}{\vec{w} \cdot \vec{w}} \vec{w}$	
Linear Independence		A set of vectors is linearly independent if no vector in the set can be written as a linear combination of the others A set of vectors $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ is linearly independent if the only solution to the the equation $c_1\vec{v}_1 + c_2\vec{v}_2 + ... + c_n\vec{v}_n = \mathbf{0}$ is $c_1 = c_2 = ... = c_n = 0$
Transformation	$T : \mathbb{R}^n \to \mathbb{R}^m$ $T(\vec{v}) = A\vec{v}$ • Additivity	• Surjective (onto) Every element in B is the image of at least one element in A. The transformation covers the entire codomain.

	$T(\vec{u}+\vec{v})=T(\vec{u})+T(\vec{v})$ • Homogeneity $T(c\vec{u})=cT(\vec{u})$	$\text{Range}(T)=B$ • Injective (one-to-one) Different inputs in A map to different outputs in B.The transformation is information-preserving — doesn't collapse distinct vectors together $T(\vec{x}_1)=T(\vec{x}_2)\Rightarrow x_1=x_2$ Or equivalently: $\ker(T)=\{\mathbf{0}\}$
Domain	$T:V\to W$ $\text{Domain}(T)=V$	Set of all **input vectors** V that the **transformation** acts on
Codomain	$T:V\to W$ $\text{Codomain}(T)=W$	Set of all possible **output vectors** W to which elements of the domain V are mapped under the **transformation**
Transpose	$\det(A)=\det(A^T)$ $(AB)^T=B^TA^T$ $(A^T)^{-1}=(A^{-1})^T$	
Image	$T:V\to W$ $\text{Im}(T)=\{\vec{w}\in W\mid\vec{w}=T(\vec{v})\text{ for some }\vec{v}\in V\}$	The image of a **transformation** $T:V\to W$ is the set of all possible outputs $T(v)$ for $v\in V$: • It is a subspace of the codomain W • If T is represented by a matrix A, the image of T is the column space of A
Preimage	$T:V\to W$ $T^{-1}(\vec{w})=\{\vec{v}\in V\mid T(\vec{v})=\vec{w}\}$	The preimage of a transformation refers to the set of all elements in the domain that map to a particular element or subset in the codomain
Span	$\text{Span}(\{v_1,v_2,...,v_k\})=\left\{\sum_{i=1}^{n}c_i\vec{v}_i\mid c_i\in\mathbb{R}\right\}$	The span of a set of vectors is the collection of all possible linear combinations of those vectors
Composition	$T_1:\mathbb{R}^n\to\mathbb{R}^m$ $\quad T_2:\mathbb{R}^m\to\mathbb{R}^p$ $T_1(\vec{v})=A\vec{v}$ $\quad T_2(\vec{v})=B\vec{v}$ $T\circ S(\vec{v})=T_2(T_1(\vec{v}))$ $T_2\circ T_1(\vec{v})=B(A\vec{v})=(BA)\vec{v}$ • Additivity $(T_3\circ T_2)\circ T_1=T_3\circ(T_2\circ T_1)$ • Homogeneity $T_2\circ T_1(c\vec{u})=c(T_2\circ T_1)(\vec{u})$ • Identity Transformation	

<table>
<tr><td></td><td>$I \circ T = T \qquad T \circ I = T$</td><td></td></tr>
<tr><td>Column Space
(Range)</td><td>$\mathrm{Col}(A) = \{Ax \mid \vec{x} \in \mathbb{R}^n\}$
Or equivalently
$A = [\vec{c}_1 \;\; \vec{c}_2 \;\; ... \;\; \vec{c}_3]$
$\mathrm{Col}(A) = \mathrm{span}(\vec{c}_1, \vec{c}_2, ..., \vec{c}_3)$</td><td>The column space (or range) of a matrix A is the set of all linear combinations of its columns</td></tr>
<tr><td>Determinant</td><td>$\det(A)$</td><td>The determinant of a square matrix A measure of the "scale factor" by which the matrix A transforms a space

$\det(A) \neq 0$

A does not collapse the space
A has full rank
A's columns are linearly independent
A is invertable

$\det(A) = 0$

A collapses the space into lower dimension
A does not have full rank
A's columns are linearly dependent
A is non-invertable (singular)

</td></tr>
<tr><td>Invertibility</td><td>$\det(A) \neq 0 \quad \Longrightarrow \text{Invertible}$
$\det(A) = 0 \quad \Longrightarrow \text{Non-Invertible}$
$AA^{-1} = A^{-1}A = I_n$
$(AB)^{-1} = B^{-1}A^{-1}$
$\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$</td><td></td></tr>
<tr><td>Basis</td><td>Linear Independence
$c_1\vec{v_1} + c_2\vec{v_2} + ... + c_k\vec{v_k} = \mathbf{0}$
$\Rightarrow c_1 = c_2 = ... = c_k = 0$
Spanning
$\forall \vec{v} \in V, \exists c_1, ..., c_k \in \mathbb{R} \quad s.t.$
$\vec{v} = c_1\vec{v_1} + ... + c_k\vec{v_k}$</td><td>
A basis of a vector space V is a set of linearly independent vectors that span the space
Every vector in V can be uniquely written as a linear combination of the basis vectors

E.g.:</td></tr>
<tr><td>Dimension</td><td>$\dim(V)$</td><td>Number of linearly independent vectors (basis) in a vector space V
$V \subseteq \mathbb{R}^n$
<table>
<tr><td>$\dim(V) = 0$</td><td>$V = \{\mathbf{0}\}$</td></tr>
<tr><td>$\dim(V) = 1$</td><td>V is a line through the origin in $\mathbb{R}^n$</td></tr>
</table></td></tr>
</table>

<table>
<tr><td></td><td></td><td>
<table>
<tr><td>$\dim(V) = 2$</td><td>V is a plane through the origin in $\mathbb{R}^n$</td></tr>
<tr><td>$\dim(V) = k$</td><td>V is a k-dimensional flat subspace of $\mathbb{R}^n$</td></tr>
<tr><td>$\dim(V) = n$</td><td>$V = \mathbb{R}^n$</td></tr>
</table>

$\mathbb{R}^2$ has dimension 2:
A basis: $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$
$\mathbb{R}^3$ has dimension 3:
A basis: $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$

</td></tr>
<tr><td>Rank</td><td>$\text{Rank}(A) = \dim(\text{Col}(A)) = \dim(\text{Row}(A))$</td><td>

The rank of a matrix A is the dimension of its column space (or row space)
Number of linearly independent columns (or rows)

</td></tr>
<tr><td>Eigen</td><td>$Ax = \lambda x, \quad x \neq \mathbf{0}$</td><td>
Set of all nonzero vectors $\vec{x}$ such that when the transformation represented by matrix A is applied to $\vec{x}$, the result is a scaled version of $\vec{x}$ itself

These vectors lie along directions that are preserved by the transformation:

$|\lambda| > 1$: stretched
$0 < |\lambda| < 1$: shrunk
$\lambda < 0$: flipped
$\lambda = 1$: stay the same

</td></tr>
<tr><td>Null Space (kernel)</td><td>$\text{Null}(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \mathbf{0}\}$</td><td>The null space of a matrix A is the set of all input vectors that get mapped to the zero vector when you multiply them by A</td></tr>
<tr><td>Identity Matrix</td><td>$$I_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$</td><td></td></tr>
<tr><td>Matrix Inverse</td><td>$A \cdot A^{-1} = I$</td><td></td></tr>
<tr><td>RREF</td><td>
1. Row Swapping (Interchange)
$$R_1 \leftrightarrow R_2$$
2. Row Scaling (Multiplication)
$$R_1 \to \frac{1}{3} R_1$$
3. Row Addition (Replacement)
$$R_1 \to R_1 - 2R_2$$
</td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

3. Matrix

$$\begin{array}{cc} & n \\ m & \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array})$$

3.1. Matrix Vector Product

$$\begin{array}{cc} & n \\ m & \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array} \qquad \begin{array}{cc} n & \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \end{array}$$

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{bmatrix}$$

3.2. Matrix Multiplication

$$A = \begin{array}{cc} & n \\ m & \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array} \qquad B = \begin{array}{cc} & p \\ n & \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{bmatrix} \end{array}$$

$$A = \begin{array}{cc} & n \\ m & \begin{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \end{bmatrix} \\ \begin{bmatrix} a_{21} & a_{22} & \dots & a_{2n} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{bmatrix} \end{array} \qquad B = \begin{array}{cc} & p \\ n & \begin{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{bmatrix} & \begin{bmatrix} b_{12} \\ b_{22} \\ \vdots \\ b_{n2} \end{bmatrix} & \dots & \begin{bmatrix} b_{1p} \\ b_{2p} \\ \vdots \\ b_{np} \end{bmatrix} \end{bmatrix} \end{array}$$

A: Row Representation

$$A = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$$

$$r_i = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix}, \quad \text{for } i = 1, 2, \dots, m$$

B: Column Representation

$$B = \begin{bmatrix} c_1 & c_2 & \dots & c_p \end{bmatrix}$$

$$c_j = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}, \quad \text{for } j = 1, 2, \dots, p$$

$$C = \begin{array}{cc} & p \\ m & \begin{bmatrix} r_1 \cdot c_1 & r_1 \cdot c_2 & \dots & r_1 \cdot c_p \\ r_2 \cdot c_1 & r_2 \cdot c_2 & \dots & r_2 \cdot c_p \\ \vdots & \vdots & \ddots & \vdots \\ r_m \cdot c_1 & r_m \cdot c_2 & \dots & r_m \cdot c_p \end{bmatrix} \end{array}$$

3.3. Transpose

$$A^T$$

$$A = \begin{array}{cc} & n \\ m & \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \end{array} \qquad A^T = \begin{array}{cc} & m \\ n & \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \end{array}$$

$$A = \begin{array}{cc} & n \\ m & \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{ij} & \cdots & a_{mn} \end{bmatrix} \end{array} \qquad A^T = \begin{array}{cc} & m \\ n & \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{i1} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{i2} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{1j} & a_{2j} & \cdots & a_{ij} & \cdots & a_{mj} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{in} & \cdots & a_{mn} \end{bmatrix} \end{array}$$

4. Matrix Factorization

4.1. LU Decomposition

$$A = LU$$

where:

- L: lower triangular matrix (entries above the diagonal are zero)
- U: upper triangular matrix (entries below the diagonal are zero)

1. $m \times n$ Matrix (with $m \geq n$)

$$L = \begin{array}{cc} & n \\ m & \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1n} \\ l_{21} & l_{22} & \cdots & l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ l_{m1} & l_{m2} & \cdots & l_{mn} \end{bmatrix} \end{array} \qquad U = \begin{array}{cc} & n \\ n & \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nn} \end{bmatrix} \end{array}$$

2. $m \times n$ Matrix (with $m < n$)

see QR Decomposition

4.2. QR Decomposition

4.3. Cholesky Decomposition

4.4. Singular Value Decomposition (SVD)

$$A = U\Sigma V^T$$

$$U = \begin{array}{cc} & m \\ m & \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & u_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mm} \end{bmatrix} \end{array} \qquad \Sigma = \begin{array}{cc} & n \\ m & \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nm} \end{bmatrix} \end{array} \qquad V^T = \begin{array}{cc} & n \\ n & \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix} \end{array}$$

	$\left[\begin{array}{ccc\|c} 1 & 2 & -1 & 2 \\ 2 & 3 & 1 & 5 \\ 3 & 4 & -2 & 4 \end{array}\right]$
$R_2 \to R_2 - 2R_1$ $R_3 \to R_3 - 3R_1$	$\left[\begin{array}{ccc\|c} 1 & 2 & -1 & 2 \\ 0 & -1 & 3 & 1 \\ 0 & -2 & 1 & -2 \end{array}\right]$
$R_2 \to -R_2$	$\left[\begin{array}{ccc\|c} 1 & 2 & -1 & 2 \\ 0 & 1 & -3 & -1 \\ 0 & -2 & 1 & -2 \end{array}\right]$
$R_1 \to R_1 - 2R_2$ $R_3 \to R_3 + 2R_2$	$\left[\begin{array}{ccc\|c} 1 & 0 & 5 & 4 \\ 0 & 1 & -3 & -1 \\ 0 & 0 & -5 & -4 \end{array}\right]$
$R_3 \to \frac{1}{-5}R_3$	$\left[\begin{array}{ccc\|c} 1 & 0 & 5 & 4 \\ 0 & 1 & -3 & -1 \\ 0 & 0 & 1 & \frac{4}{5} \end{array}\right]$
$R_1 \to R_1 - 5R_3$ $R_2 \to R_2 + 3R_3$	$\left[\begin{array}{ccc\|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{7}{5} \\ 0 & 0 & 1 & \frac{4}{5} \end{array}\right]$

$$A = \begin{bmatrix} 2 & -1 & -3 \\ -4 & 2 & 6 \end{bmatrix}$$

1. Null Space

The **null-space** of A, denoted $N(A)$, consists of all vectors $\vec{x} \in \mathbb{R}^3$ such that $Ax = 0$. The set of all such vectors is the **pre-image** of the zero vector under the transformation defined by A. In other words, $N(A) = \{x \in \mathbb{R}^3 \mid A\vec{x} = 0\}$, which represents the set of vectors that A maps to zero.

$$N(A) = \{\vec{x} \in \mathbb{R}^3 \mid A\vec{x} = \mathbf{0}\}$$

To find the null space $N(A)$ of the matrix A, we can use the **row-reduced echelon form (RREF)**. By augmenting the matrix A with a zero column and performing row operations, we reduce it to the form:

$$\begin{bmatrix} 2 & -1 & -3 \\ -4 & 2 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

	$\left[\begin{array}{ccc\|c} 2 & -1 & -3 & 0 \\ -4 & 2 & 6 & 0 \end{array}\right]$

$R_1 \to \frac{R_1}{2}$ $R_2 \to \frac{R_2}{4}$	$\left[\begin{array}{ccc\|c} 1 & -\frac{1}{2} & -\frac{3}{2} & 0 \\ -1 & \frac{1}{2} & \frac{3}{2} & 0 \end{array}\right]$
$R_2 \to R_2 - R_1$	$\left[\begin{array}{ccc\|c} 1 & -\frac{1}{2} & -\frac{3}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$

$$\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{3}{2} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$x_1 - \frac{1}{2}x_2 - \frac{3}{2}x_2 = 0$$
$$x_1 = \frac{1}{2}x_2 + \frac{3}{2}x_2$$
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} \frac{1}{2} \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \frac{3}{2} \\ 0 \\ 1 \end{bmatrix}$$
$$N(a) = \operatorname{span}\left(\left\{ \begin{bmatrix} \frac{1}{2} \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} \frac{3}{2} \\ 0 \\ 1 \end{bmatrix} \right\}\right)$$

The dimension of the **null-space** is the number of vectors in this basis, which is 2. This is important because the dimension of the null space gives us insight into how many degrees of freedom exist in the system of equations $Ax = 0$

2. Column Space

$$C(A) = \operatorname{span}\left(\left\{ \begin{bmatrix} 2 \\ -4 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \begin{bmatrix} -3 \\ 6 \end{bmatrix} \right\}\right)$$
$$= \operatorname{span}\left(\left\{ \begin{bmatrix} 2 \\ -4 \end{bmatrix} \right\}\right)$$

3. Basis

$$\begin{bmatrix} 2 \\ -4 \end{bmatrix}$$

4. Rank

Number of vector in the basis of our column space

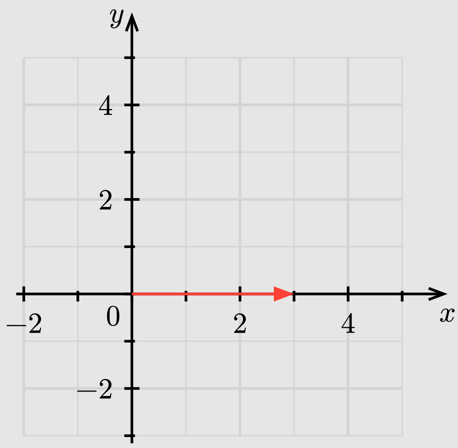
$$\operatorname{Rank}(A) = 1$$

5. Vectors

$$\text{Vector} = \text{Magnitude} + \text{Direction}$$

A car is moving:

$$\underbrace{\underbrace{\text{3 MPH}}_{\substack{\text{Magnitude}\\ \text{(Speed} \to \text{Scalar)}}} \qquad \underbrace{\text{East}}_{\text{Direction}}}_{\text{Velocity (Vector)}}$$



$$\vec{v} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

5.1. Real Coordinate Spaces

N-dimensional Real Coordinate Space

$$\mathbb{R}^n$$

$$\vec{x} \in \mathbb{R}^n$$

All possible real-valued n-tuples

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

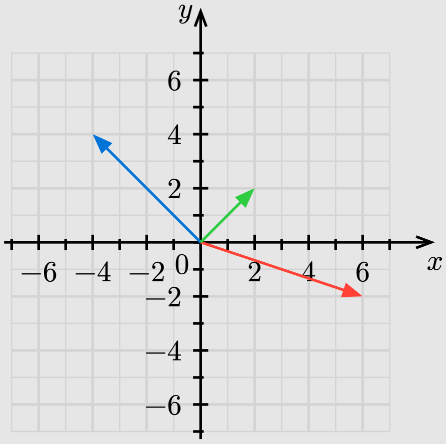
5.2. Vector Operations

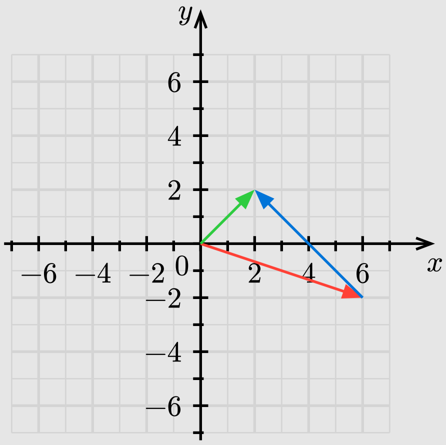
5.2.1. Vector Addition

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix}$$

$$\vec{a} = \begin{bmatrix} 6 \\ -2 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$$

$$\vec{a} + \vec{b} = \begin{bmatrix} 6 + -4 \\ -2 + 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$



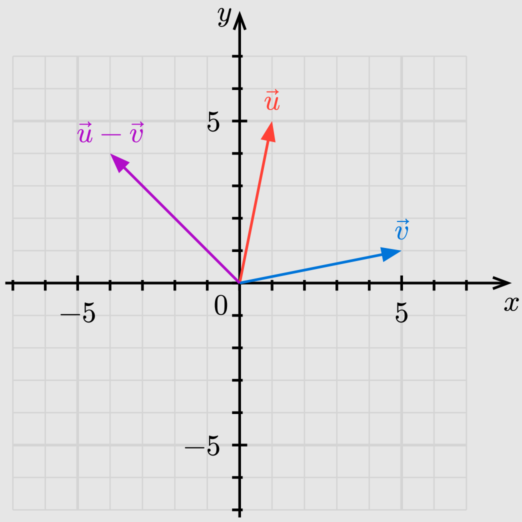


5.2.2. Vector Subtraction

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \\ x_3 - y_3 \end{bmatrix}$$

$$\vec{u} = \begin{bmatrix} 1 \\ 5 \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

$$\vec{u} - \vec{v} = \begin{bmatrix} 1 - 5 \\ 5 - 1 \end{bmatrix} = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$$

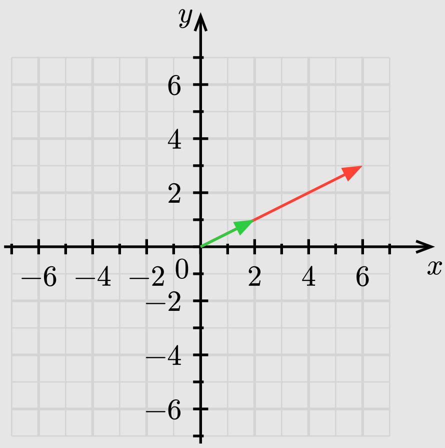


5.2.3. Scalar Multiplication

$$c \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} c \times x_1 \\ c \times x_2 \\ c \times x_3 \end{bmatrix}$$

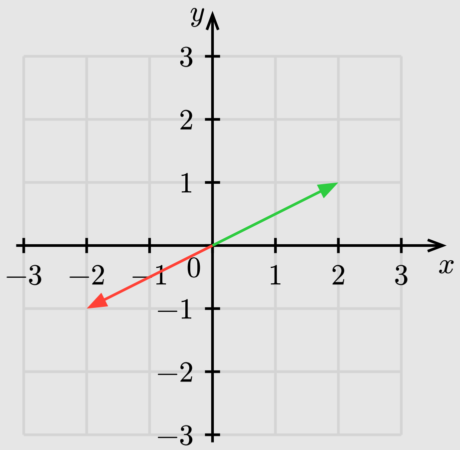
$$\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$3\vec{a} = 3 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 \\ 3 \cdot 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$$



$$\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$-1\vec{a} = -1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \cdot 2 \\ -1 \cdot 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$



5.3. Unit Vector

A vector that has a magnitude (or length) of exactly 1

For a vector $\vec{v}$ in n-dimensional space, a unit vector $\hat{v}$ is defined as:

$$\hat{v} = \frac{\vec{v}}{\| \vec{v} \|}$$

Where:

- $\|\vec{v}\|$ is the **magnitude** (or **norm**) of the vector $\vec{v}$, computed as:

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

Key Properties

- **Magnitude**:

$$\|\hat{v}\| = 1$$

- **Direction**: A unit vector points in the same direction as the original vector $\vec{v}$

Finding unit vector (vector of magnitude 1) with given direction

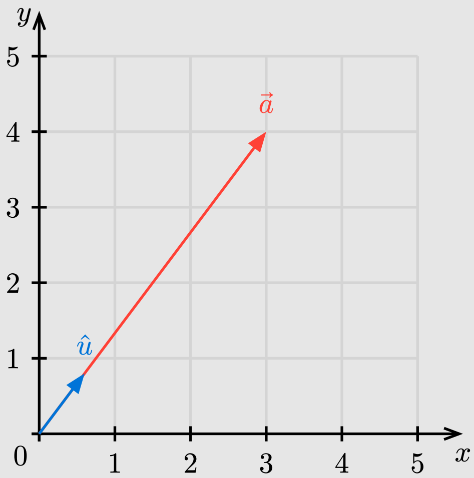
$$\vec{a} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

Magnitude

$$\| \vec{a} \| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$$

$$\hat{u} = \left(\frac{3}{\| \vec{a} \|}, \frac{4}{\| \vec{a} \|} \right) = \left(\frac{3}{5}, \frac{4}{5} \right)$$

$$\| \hat{u} \| = \sqrt{\left(\frac{3}{5}\right)^2 + \left(\frac{4}{5}\right)^2} = \sqrt{\frac{9}{25} + \frac{16}{25}} = \sqrt{\frac{25}{25}} = \sqrt{1} = 1$$



5.4. Parametric Representation of line

Set L of all points (i.e., line) equal to the set of all vectors $\vec{x}$ plus some scalar t times the vector $\vec{v}$ such that t can be any real number ($\mathbb{R}$)

$$L = \{\vec{x} + t\vec{v} \mid t \in \mathbb{R}\}$$

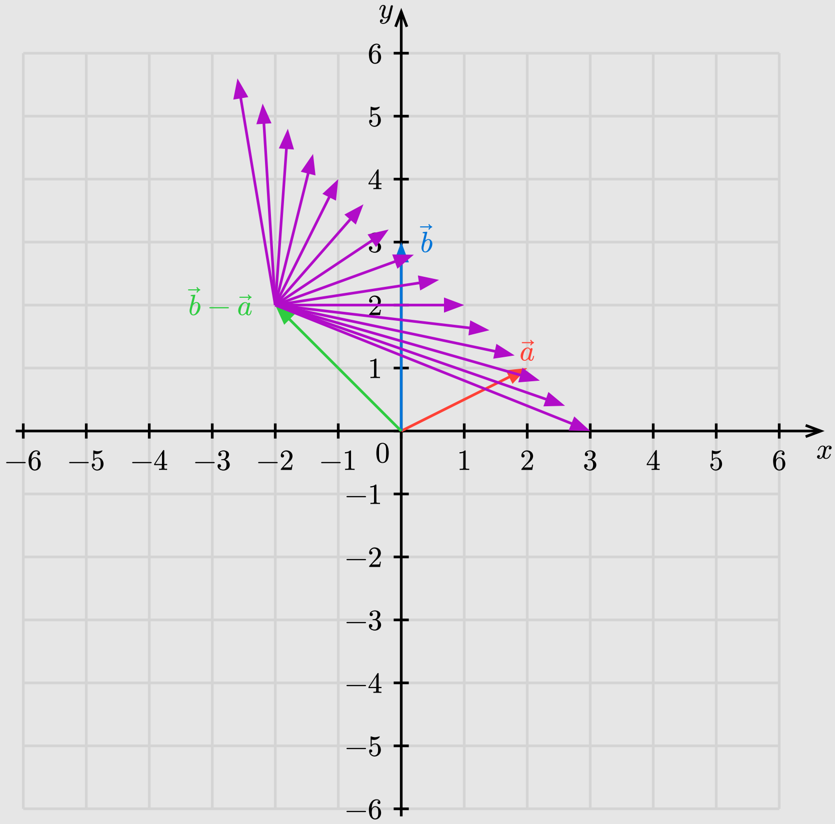
$$\vec{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$t = 1$$

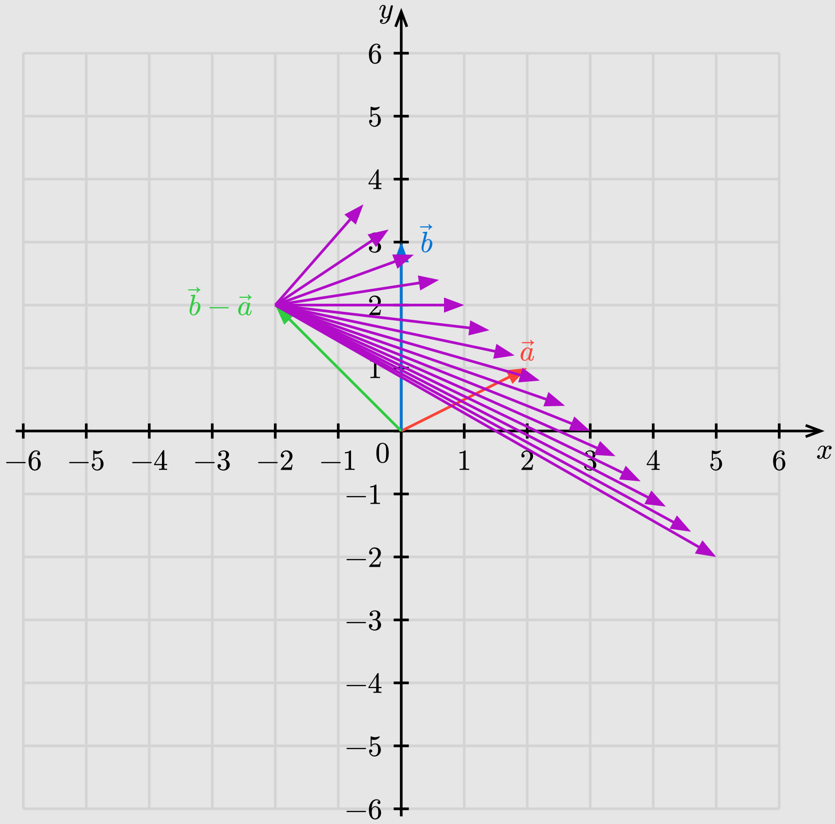
The line L can be defined as:

$$\begin{aligned} \vec{b} + t(\vec{b} - \vec{a}) &= \begin{bmatrix} 0 \\ 3 \end{bmatrix} + 1\left(\begin{bmatrix} 0 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right) \\ &= \begin{bmatrix} 0 \\ 3 \end{bmatrix} + 1\begin{bmatrix} -2 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} -2 \\ 5 \end{bmatrix} \end{aligned}$$



The line L can also be defined as:

$$\begin{aligned}\vec{a}+t\left(\vec{b}-\vec{a}\right) &= \begin{bmatrix}2\\1\end{bmatrix}+1\left(\begin{bmatrix}0\\3\end{bmatrix}-\begin{bmatrix}2\\1\end{bmatrix}\right)\\ &= \begin{bmatrix}2\\1\end{bmatrix}+1\begin{bmatrix}-2\\2\end{bmatrix}\\ &= \begin{bmatrix}0\\3\end{bmatrix}\end{aligned}$$



Generalization

$$P_1 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad P_2 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$L = P_1 + t(P_1 - P_2) \mid t \in \mathbb{R}$$

$$\overrightarrow{P_1} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \qquad \overrightarrow{P_2} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$\overrightarrow{P_1} - \overrightarrow{P_2} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

- L starts at P_1 and moves toward P_2 as t decreases

$$L = P_1 + t(P_1 - P_2) = \begin{bmatrix} -1 \\ 2 \end{bmatrix} + t \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$x = -1 + t(-1)$$
$$y = 2 + t(-1)$$

- L starts at P_2 and moves toward P_1 as t increases

$$L = P_2 + t(P_1 - P_2) = \begin{bmatrix} 0 \\ 3 \end{bmatrix} + t \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$x = 0 + t(-1)$$
$$y = 3 + t(-1)$$

- L starts at P_1 and moves toward P_2 as t increases

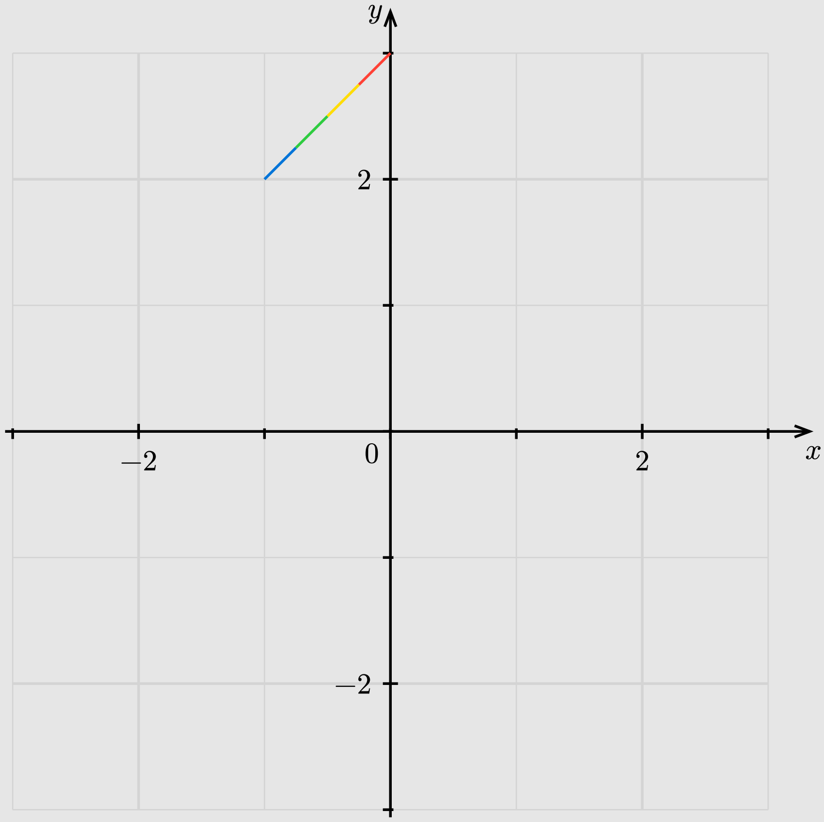
$$L = P_1 + t(P_2 - P_1) = \begin{bmatrix} -1 \\ 2 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$x = -1 + t(1)$$
$$y = 2 + t(1)$$

- L starts at P_2 and moves toward P_1 as t decreases

$$L = P_2 + t(P_2 - P_1) = \begin{bmatrix} 0 \\ 3 \end{bmatrix} + t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$x = 0 + t(1)$$
$$y = 3 + t(1)$$



5.5. Vector Spaces

Let's say your factory can produce up to 300 units of product 1, 500 units of product 2, and 400 units of product 3. The set of all possible production combinations forms a vector space:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$0 \leq x_1 \leq 300$$

$$0 \leq x_2 \leq 500$$

$$0 \leq x_3 \leq 400$$

6. Matrices

$m \times n$ matrix $\boldsymbol{A}$

- m: rows
- n: columns

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

6.1. Matrix-Vector Products

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\boldsymbol{A}\vec{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

For the dot product to be defined, the number of columns in the matrix $\boldsymbol{A}$ (which is n) must match the number of elements in the vector $\vec{x}$ (also n).

The result of multiplying matrix $\boldsymbol{A}$ and vector $\vec{x}$ will be a column vector with dimensions $m \times 1$, where m is the number of rows in the matrix $\boldsymbol{A}$

$$(m \times n) \cdot (n \times 1) = m \times 1$$

1. As Row vectors

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\vec{a}^T = [a_1, a_2, ..., a_n]$$
$$\vec{b}^T = [b_1, b_2, ..., b_n]$$

$$\boldsymbol{A} = \begin{bmatrix} [a_1, a_2, ..., a_n] \\ [b_1, b_2, ..., b_n] \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} \vec{a} \\ \vec{b} \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\begin{bmatrix} \vec{a}^T \\ \vec{b}^T \end{bmatrix} \cdot \vec{x} = \begin{bmatrix} \vec{a} \cdot \vec{x} \\ \vec{b} \cdot \vec{x} \end{bmatrix}$$

2. As Column Vectors

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} & \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} \vec{a} & \vec{b} \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\boldsymbol{A}\vec{x} = x_1\vec{a} + x_2\vec{b}$$

6.2. Null Space

The null space (or kernel) of a matrix $\boldsymbol{A}$ is the set of all vectors $\boldsymbol{x}$ that satisfy the equation:

$$\boldsymbol{A}\vec{x} = \mathbf{0}$$

Where:

- $\boldsymbol{A}$: $m \times n$ matrix
- $\vec{x}$: n-dimansional vector
- $\mathbf{0}$: zero vecor in $\mathbb{R}^m$

$$N(\boldsymbol{A}) = N(\operatorname{rref}(\boldsymbol{A})) = \operatorname{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$

We want to find the null space of A, which consists of all vectors $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ that satisfy:

$$\boldsymbol{A}\vec{x} = \mathbf{0}$$

This expands to the following system of linear equations:

$$\begin{cases} 1x_1 + 1x_2 + 1x_3 + 1x_4 = 0 \\ 1x_1 + 2x_2 + 3x_3 + 4x_4 = 0 \\ 4x_1 + 3x_2 + 2x_3 + 1x_4 = 0 \end{cases}$$

This can be represented as the augmented matrix:

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 0 \\ 4 & 3 & 2 & 1 & 0 \end{array}\right]$$

6.2.1. Column Space

The **columns space** (or range) of matrix A is span of its columns vectors

If the matrix A has columns $\vec{a}_1, \vec{a}_2, ..., \vec{a}_n$, then the column space of A is defined as:

$$\operatorname{Col}(A) = \{\vec{y} \in \mathbb{R}^m \mid \vec{y} = A\vec{x} \text{ for some } \vec{x} \in \mathbb{R}^n\}$$

or equivalently,

$$\operatorname{Col}(A) = \operatorname{span}(\{\vec{a}_1, \vec{a}_2, ..., \vec{a}_n\})$$

Consider the simple example of a 2×2 matrix:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$$

The matrix has two columns:

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{a}_2 = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$$

The column space, denoted $\operatorname{Col}(A)$, is the span of these two vectors:

$$\mathrm{Col}(A) = \mathrm{span}\left(\left\{\begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}2\\6\end{bmatrix}\right\}\right)$$

Finding the Column Space

We observe that the two columns $\vec{a}_1$ and $\vec{a}_2$ are **linearly dependent**:

$$\vec{a}_2 = k\vec{a}_1$$

This means that $\vec{a}_2$ is a scalar multiple of $\vec{a}_1$, the the two columns are **linearly dependent**. As a result, the column space is spanned by just one vector, $\vec{a}_1$, because any linear combination of $\vec{a}_1$ and $\vec{a}_2$ can be reduced to a multiple of $\vec{a}_1$.

Therefore, the column space of A is:

$$\mathrm{Col}(A) = \mathrm{span}\left(\left\{\begin{bmatrix}1\\3\end{bmatrix}\right\}\right)$$

which represents all vectors of the form:

$$c\begin{bmatrix}1\\3\end{bmatrix} = \begin{bmatrix}c\\3c\end{bmatrix} \qquad \text{for any scalar } c$$

In other words, the column space is a line in $\mathbb{R}^2$ through the origin in the direction of

Rank of A

The rank of A, which is the **dimension of its column space**, is 1 because there is only one linearly independent column

This means the column space is the span of the columns of A, or all vectors that can be formed by taking linear combinations of the columns of A.

6.2.2. Dimension of a Subspace

Number of elements in a basis for the subspace

6.2.3. Nullity

Dimension of the **Null Space**

$$\dim(N(\boldsymbol{A}))$$

The nullity of A: number of non-pivot columns (i.e., free variables) in the rref of A

6.2.4. Rank

Dimension of the column **space**

$$\mathrm{rank}(A) = \dim(C(A))$$

6.2.5. Matrix Representation of Systems of Equations

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + ... + a_{1m}x_m = b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2m}x_m = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + ... + a_{nm}x_m = b_n \end{gathered}$$

Coefficient Matrix (A):

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

Variable Vector (x):

$$\mathrm{x} \ = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

Constant Vector (b):

$$\mathrm{b} \ = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\mathrm{Ax} = b$$

The system of equations:

$$\begin{aligned} 2x_1 + 3x_2 + 5x_3 &= 100 \\ 4x_1 + 2x_2 + 1x_3 &= 80 \\ 1x_1 + 5x_2 + 2x_3 &= 60 \end{aligned}$$

Can be represented as a matrix equation:

$$\begin{bmatrix} 2 & 3 & 5 \\ 4 & 2 & 1 \\ 1 & 5 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 100 \\ 80 \\ 60 \end{bmatrix}$$

6.3. Matrix Multiplication

$m \times n$ matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

$n \times p$ matrix:

$$B = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix}$$

Compute Each Element of Result Matrix C

$$c_{ij} = \sum_{k=1}^{n} = a_{ik} b_{kj}$$

Let A be an $n \times m$ matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Let B an $p \times n$ matrix:

$$B = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}$$

Calculate Each Element of C

$$\begin{aligned} c_{11} &= (1 \cdot 7) + (2 \cdot 9) + (3 \cdot 11) = 58 \\ c_{12} &= (1 \cdot 8) + (2 \cdot 10) + (3 \cdot 12) = 64 \\ c_{21} &= (4 \cdot 7) + (5 \cdot 9) + (6 \cdot 11) = 138 \\ c_{22} &= (4 \cdot 8) + (5 \cdot 10) + (6 \cdot 12) = 154 \end{aligned}$$

C is a $m \times p$ matrix

$$C = \begin{bmatrix} 58 & 64 \\ 139 & 154 \end{bmatrix}$$

7. Linear Combinations

Set of vector

$$v_1, v_2, ..., v_n \in \mathbb{R}^n$$

Where

- $v_1, v_2, ..., v_n$: set of vectors
- $\mathbb{R}^n$: set of all ordered tuples of n real numbers

Linear combination of those vector

$$c_1 v_1 + c_2 v_2 + ... + c_n v_n$$

$$c_1, c_2, ..., c_n \in \mathbb{R}$$

Where:

- $c_1, c_2, ..., c_n$: constants or weights

$$\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$0\vec{a} + 0\vec{b}$$

$$3\vec{a} + 2\vec{b}$$

8. Span

Represents the subspace of the vector space that is “covered” by these vectors through their linear combinations

If you have a set of vectors $v_1, v_2, ..., v_n$, the span of these vectors is the set of all vectors that can be written as:

$$\mathrm{Span}(v_1, v_2, ..., v_n) = \{c_1 v_1 + c_2 v_2 + ... + c_n v_n \mid c_1, c_2, ..., c_n \in \mathbb{R}\}$$

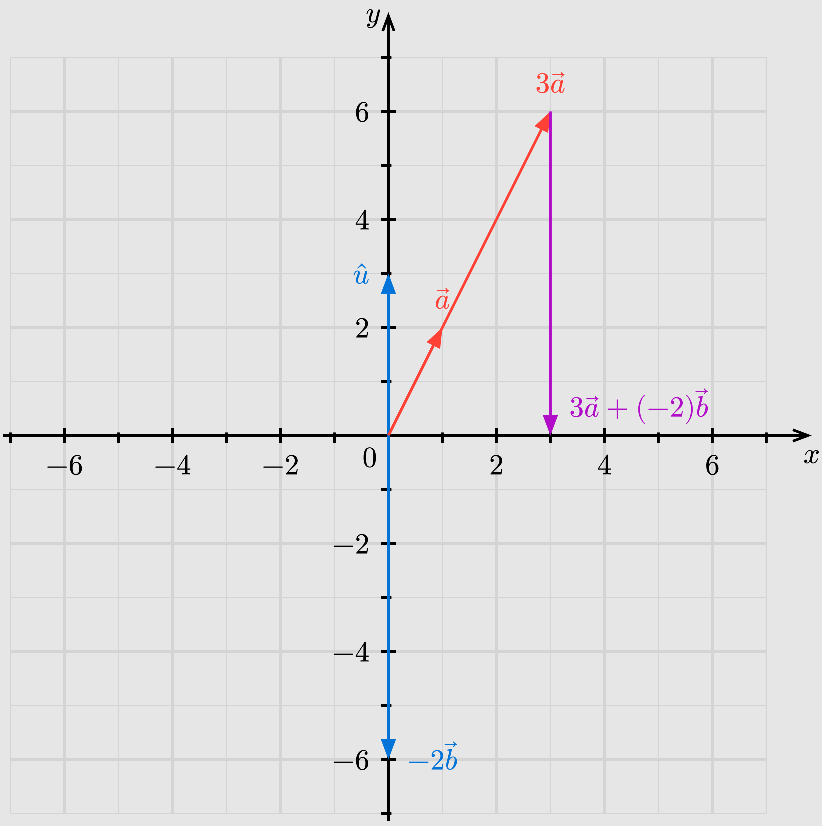
Any vector in $\mathbb{R}^2$ can be represented by a linear combination with some combination of these vectors

1. Spanning $\mathbb{R}^2$

$$\mathrm{Span}\left(\vec{a}, \vec{b}\right) = \mathbb{R}^2$$

$$\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$3\vec{a} + (-2)\vec{b} = \begin{bmatrix} 3 - 0 \\ 6 - 6 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$



Any point $\vec{x}$ can be represented as a linear combination of $\vec{a}$ and $\vec{b}$

1. Define the vectors

$$\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} 0 \\ 3 \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

2. Express $\vec{x}$ as a linear combinations

$$c_1 \vec{a} + c_2 \vec{b} = \vec{x}$$

Which expands to

$$c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

3. Set up the system of equations

$$1c_1 + 0c_2 = x_1 \qquad (1)$$
$$2c_1 + 3c_2 = x_2 \qquad (2)$$

4. Express c_1: From equation (1), we can directly express c_1

$$c_1 = x_1$$

5. Substitute c_1 into equation (2)

$$2x_1 + 3c_2 = x_2$$

Rearranging gives:

$$3c_2 = x_2 - 2x_1$$

6. Solve for c_2 : Dividing both sides by 3 yields

$$c_2 = \frac{x_2 - 2x_1}{3}$$

7. Example with specific values: Let's say we want to find c_1 and c_2 when $\vec{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

Substitute $x_1 = 2$ and $x_2 = 2$

$$c_1 = x_1 = 2$$
$$c_2 = \frac{2 - 2 \cdot 2}{3} = -\frac{2}{3}$$

8. Final linear combination: Now, substituting c_1 and c_2 back into the linear combination

$$2\vec{a} - \frac{2}{3}\vec{b} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Verifying

$$2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \frac{1}{3} \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

9. This shows that

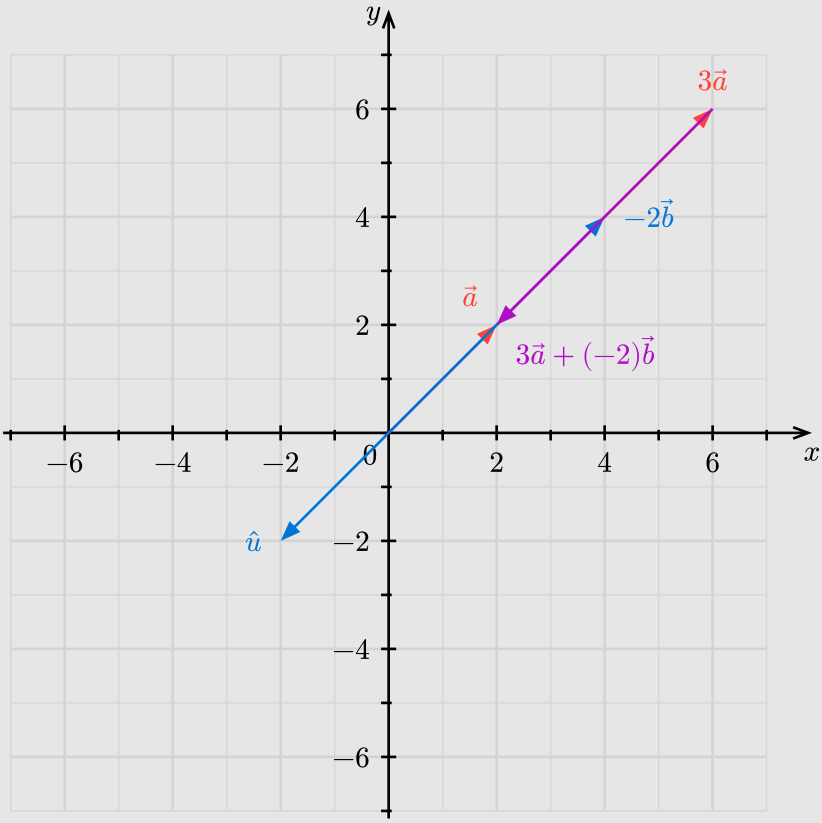
$$2\vec{a} - \frac{2}{3}\vec{b} = \vec{x}$$

2. Spanning Line in $\mathbb{R}^2$

Any linear combination of $\vec{a}$ and $\vec{b}$ will produce vectors that lie along the same line. This is the line through the origin in the direction of $\vec{a}$ (or $\vec{b}$), with all points on the line being scalar multiples of $\vec{a}$

$$\vec{a} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

$$3\vec{a} + (-2)\vec{b} = \begin{bmatrix} 6-4 \\ 6-4 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$



9. Linear Independence

1. Definition of Linear Independence

The set of vectors

$$S = \{v_1, v_2, \ldots, v_n\}$$

is said to be **linearly independent** if the only solution to the equation

$$c_1 v_1 + c_2 v_2 + \ldots + c_n v_n = \mathbf{0}$$

is $c_1 = c_2 = \ldots = c_n = 0$. In other words, no vector in the set can be written as a linear combination of the others.

If at least one constant c_i is non-zero, the set is linearly dependent.

Example 1: Testing for Linear Independence

Problem: Is the following set of vectors **linearly dependent**?

$$S = \{\vec{v}_1 \vec{v}_2\}$$

Where:

$$\vec{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{and} \quad \vec{v}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

For a set of vectors to be **linearly independent**, the only solution to the equation:

$$c_1\vec{v}_1 + c_2\vec{v}_2 = \mathbf{0}$$

must be $c_1 = 0$ and $c_2 = 0$

In this case:

$$c_1\begin{bmatrix}2\\1\end{bmatrix} + c_2\begin{bmatrix}3\\2\end{bmatrix} = \mathbf{0}$$

If not only the zero solution exists (i.e., if c_1 or c_2 can be non-zero), the set is **linearly dependent**.

Step 1. Set up the system of equations:

1. $2c_1 + 3c_2 = 0$
2. $1c_1 + 2c_2 = 0$

2. Eliminate one variable

$$2 \times (1c_1 + 2c_2 = 0) \Rightarrow 2c_1 + 4c_2 = 0$$

Now the system is

$$2c_1 + 3c_2 = 0$$
$$2c_1 + 4c_2 = 0$$

3. Subtract the equations

$$(2c_1 + 3c_2) - (2c_1 + 4c_2) = 0$$

Simplifies to:

$$(2c_1 - 2c_1) + (4c_2 - 3c_2) = 0$$

So:

$$c_2 = 0$$

4. Substitute back to find c_1

Now that we know $c_2 = 0$, substitute this value into one of the original equations. Let's use the second equation:

$$1c_1 + 2c_2 = 0$$

Substitute $c_2 = 0$:

$$1c_1 + 2(0) = 0$$
$$c_1 = 0$$

Conclusion:

Since $c_1 = 0$ $c_2 = 0$, the set of vectors S is **linearly independent**. These vecots span $\mathbb{R}^2$.

Example 2: Testing for Linear Dependence

Problem: Is the following set of vectors **linearly dependent**?

$$S = \{\vec{v}_1 \vec{v}_2\}$$

Where:

$$\vec{v}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{v}_2 = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

The span of this set is the collection of all vectors that can be formed by linear combinations of $\vec{v}_1$ and $\vec{v}_2$:

$$c_1 v_1 + c_2 v_2$$

Since $v_2 = 2v_1$, the linear combination becomes:

$$c_1 v_1 + c_2 (2v_1) = (c_1 + 2c_2) v_1$$

$$c_1 \begin{bmatrix} 2 \\ 3 \end{bmatrix} + c_2 \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

$$c_1 \begin{bmatrix} 2 \\ 3 \end{bmatrix} + c_2 2 \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$(c_1 + 2c_2) \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$c_3 \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Thus, any linear combination of these vectors is just a scalar multiple of v_1. The span is a single line in $\mathbb{R}^2$, and the vectors are **linearly dependent**.

For any two **colinear** vectors in $\mathbb{R}^2$, their span reduces to a single line.

One vector in the set can be represented by some combination of other vectors in the set

2. General Rule

In $\mathbb{R}^n$, if you have more than n vectors, at least one vector must be linearly dependent on the others, meaning the set cannot be linearly independent.

10. Subspace

V is a linear subspace of $\mathbb{R}^n$:

- **Non-emptiness**: V contains the **0** vector

$$\mathbf{0} \in V$$

- **Closure under addition**: If u and v are any vectors in the subspace V, then their sum $u + v$ must also be in V.

$$\text{If } u, v \in V, \text{then } u + v \in V$$

- **Closure under scalar multiplication**: If u is any vector in V and c is any scalar (real number), then the product cu must also be in V.

$$\text{If } u \in V \text{ and } c \in V, \text{then } cu \in V$$

Example 1: Subspace

Problem: Is V a subspace of $\mathbb{R}^2$

$$V = \{\mathbf{0}\} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$$

- **Non-emptiness** ✅

$$\mathbf{0} \in V$$

- **Closure under addition** ✅

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- **Closure under scalar multiplication** ✅

$$c \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

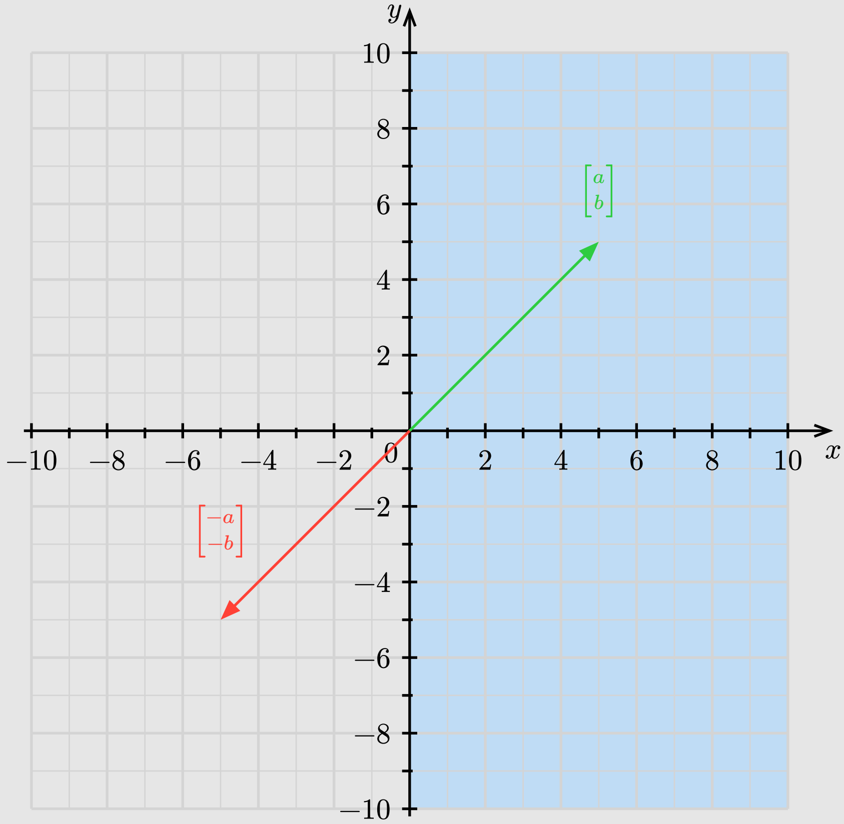
Conclusion

The subset V of $\mathbb{R}^3$ is a **subspace**

Example 2: Not Subspace

Problem: Is S a subspace of $\mathbb{R}^2$ ✅

$$S = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 \mid x_1 \geq 0 \right\}$$



- **Non-emptiness** ✅

$$\mathbf{0} \in S$$

- **Closure under addition** ✅

$$\begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} a + c \\ b + d \end{bmatrix}$$

$$\begin{aligned} a &\geq 0 \\ b &\geq 0 \\ a + b &\geq 0 \end{aligned}$$

- **Closure under scalar multiplication** ❌

$$-1 \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -a \\ -b \end{bmatrix}$$

Conclusion

Span and Subspace

The span of any set of vectors is a valid subspace

$$U = \text{Span}(v_1, v_2, ..., v_n) = \text{Valid Subspace of } \mathbb{R}^n$$

- **Non-emptiness**

$$0v_1 + 0v_2 + ... + 0v_n = \mathbf{0}$$

- **Closure under addition**

$$\vec{X} = a_1 v_1 + a_2 v_2 + \dots + a_n v_n$$
$$\vec{Y} = b_1 v_1 + b_2 v_2 + \dots + b_n v_n$$

$$\begin{aligned} \vec{X} + \vec{Y} &= (a_1 + b_1) v_1 + (a_2 + b_2) v_2 + \dots + (a_n + b_n) v_n \\ &= c_1 v_1 + c_2 v_2 + \dots + c_n v_n \end{aligned}$$

- **Closure under scalar multiplication**

$$\vec{X} = a_1 v_1 + a_2 v_2 + \dots + a_n v_n$$

$$\begin{aligned} b\vec{X} &= bc_1 v_1 + bc_2 v_2 + \dots + bc_n v_n \\ &= c_1 v_1 + c_2 v_2 + \dots + c_n v_n \end{aligned}$$

11. Basis

Non-redundant set of vectors that span $\mathbb{R}^n$

A basis of a vector space is a set of vectors that satisfies two conditions:

1. **Linear Independence**: No vector in the set can be written as a linear combination of the others. This means that **the only way to combine the vectors to get the zero vector is by using all zero coefficients**.

   The vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are linearly independent if the only solution to the equation

   $$c_1 \vec{v}_1 + c_2 \vec{v}_2 + \dots + c_n \vec{v}_n = \mathbf{0}$$

   is $c_1 = c_2 = \dots = c_n = 0$, where c_i are scalar coefficients.

2. **Spanning**: The set of vectors can be linearly combined to form any vector in the vector space. In other words, **every vector in the vector space can be expressed as a linear combination of the basis vectors**.

   The set $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ spans the vector space V if any vector $v \in V$ can be expressed as a linear combination of the basis vectors:

   $$\vec{v} = c_1 \vec{v}_1 + c_2 \vec{v}_2 + \dots + c_n \vec{v}_n$$

   for some scalars $c_1, c_2, \dots, c_n$.

Consider the vector space $\mathbb{R}^2$ (the 2-dimensional Euclidean space). A common basis for $\mathbb{R}^2$ is $\{e_1, e_2\}$, where:

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

This set is a basis because:

- Linear Independence: The only solution to $c_1 e_1 + c_2 e_2 = \mathbf{0}$ is $c_1 = c_2 = 0$.
- Spanning: Any vector $\vec{v} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ can be written as $\vec{v} = x e_1 + x e_2$

This means $\{e_1, e_2\}$ is a basis for $\mathbb{R}^2$, and the dimension of $\mathbb{R}^2$ is 2.

11.1. Vector Dot Product

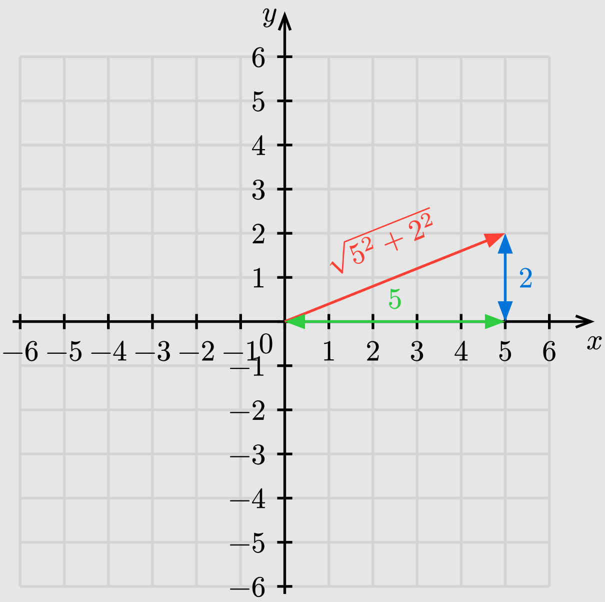
$$\vec{a} \cdot \vec{b} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n$$

11.1.1. Magnitude (Length)

$$\|\vec{a}\| = \sqrt{a_1^2 + a_2^2 + \ldots + a_n^2}$$

$$\vec{a} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

$$\|\vec{a}\| = \sqrt{5^2 + 2^2}$$



$$\vec{a} \cdot \vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a_1^2 + a_2^2 + \ldots + a_n^2$$

$$\|\vec{a}\| = \sqrt{\vec{a} \cdot \vec{a}}$$

$$\|\vec{a}\|^2 = \vec{a} \cdot \vec{a}$$

11.1.2. Properties

11.1.2.1. Commutative

$$\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$$

11.1.2.2. Distributive

$$(\vec{v} + \vec{w}) \cdot \vec{x} = \vec{v} \cdot \vec{x} + \vec{w} \cdot \vec{x}$$

11.1.2.3. Associativity

$$(c\vec{v}) \cdot \vec{w} = c(\vec{v} \cdot \vec{w})$$

11.1.3. Cauchy-Schwarz Inequality

$$|\vec{u} \cdot \vec{v}| \leq \|\vec{u}\| \ \|\vec{v}\|$$
$$|\vec{u} \cdot \vec{v}| = \|\vec{u}\| \ \|\vec{v}\| \quad \text{when} \quad \vec{u} = c\vec{v}$$

Where:

- $\vec{u} \cdot \vec{v}$: dot product of vectors $\vec{u}$ and $\vec{v}$
- $\|\vec{u}\|$ and $\|\vec{v}\|$: magnitudes (lengths) of vectors $\vec{u}$ and $\vec{v}$

Step 1: Understand the dot product

The dot product of two vectors $\vec{u} = [u_1, u_2, ..., u_n]$ and $\vec{v} = [v_1, v_2, ..., v_n]$ is calculated as:

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + ... + u_n v_n$$

The magnitude (or norm) of a vector $\vec{u}$ is:

$$\|\vec{u}\| = \sqrt{u_1^2 + u_2^2 + ... + u_n^2}$$

Step 2: Define a new function

We introduce a parameter $t \in \mathbb{R}$ and define a new vector:

$$w(t) = u - tv$$

Now, consider the dot product of this new vector with itself, which is always non-negative because it represents the square of the magnitude of $w(t)$:

$$w(t) \cdot w(t) \geq 0$$

This inequality makes sense because the dot product of any vector with itself is the square of its magnitude, and **a square is always non-negative**.

Step 3: Expand the dot product

Expand $w(t) \cdot w(t)$:

$$w(t) \cdot w(t) = (\vec{u} - t\vec{v}) \cdot (\vec{u} - t\vec{v})$$

Now, we apply the distributive property of the dot product, which behaves similarly to the distributive property of multiplication. We expand each term:

$$(\vec{u} - t\vec{v}) \cdot (\vec{u} - t\vec{v}) = \vec{u} \cdot \vec{u} - t(\vec{u} \cdot \vec{v}) - t(\vec{v} \cdot \vec{u}) + t^2(\vec{v} \cdot \vec{v})$$

Since the dot product is commutative ($\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$), we can rewrite this as:

$$\vec{u} \cdot \vec{u} - 2t(\vec{u} \cdot \vec{v}) + t^2(\vec{v} \cdot \vec{v})$$

This simplifies to:

$$\|\vec{u}\|^2 - 2t(\vec{u} \cdot \vec{v}) + t^2 \ \|\vec{v}\|^2$$

We've now expressed the result of expanding the dot product as a quadratic expression in t, where

- $\|\vec{u}\|^2$ is a constant term,
- $-2t(\vec{u} \cdot \vec{v})$ is the linear term in t
- $t^2 \ \|\vec{v}\|^2$ is the quadratic term

Step 4: Treat as a quadratic equation

Now that we have the quadratic expression:

$$\|\vec{v}\|^2\ t^2 - 2(\vec{u} \cdot \vec{v})t + \|\vec{u}\|^2 \geq 0$$

We recognize this as a standard quadratic inequality of the form $at^2 + bt + c \geq 0$, where:

- $a = \|\vec{v}\|^2$
- $b = -2(\vec{u} \cdot \vec{v})$
- $c = \|\vec{u}\|^2$

For any quadratic expression $at^2 + bt + c$ to always be non-negative, its discriminant must be less than or equal to zero. The discriminant of a quadratic equation $at^2 + bt + c = 0$ is given by:

$$\Delta = b^2 - 4ac$$

Substituting in the values of a, b, and c from our expression:

$$\Delta = (-2(\vec{u} \cdot \vec{v}))^2 - 4 \cdot \|\vec{v}\|^2 \cdot \|\vec{u}\|^2$$

Simplifying:

$$\Delta = 4(\vec{u} \cdot \vec{v})^2 - 4\ \|\vec{v}\|^2\ \|\vec{u}\|^2$$

Step 5: Apply the discriminant condition

For the quadratic inequality to hold, the discriminant must be less than or equal to zero:

$$\Delta = 4(\vec{u} \cdot \vec{v})^2 - 4\ \|\vec{v}\|^2\ \|\vec{u}\|^2 \leq 0$$

Divide by 4:

$$\Delta = (\vec{u} \cdot \vec{v})^2 \leq \|\vec{v}\|^2\ \|\vec{u}\|^2$$

Take the square root of both sides:

$$|\vec{u} \cdot \vec{v}| \leq \|\vec{u}\|\ \|\vec{v}\|$$

$$\vec{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

Step 1: Compute the dot product $\vec{u} \cdot \vec{v}$

$$\vec{u} \cdot \vec{v} = (1)(3) + (2)(4) = 3 + 8 = 11$$

Step 2: Compute the norms of $\vec{u}$ and $\vec{v}$

- The norm $\|\vec{u}\|$ is:

$$\|\vec{u}\| = \sqrt{1^2 + 2^2} = \sqrt{1 + 4} = \sqrt{5}$$

- The norm $\|\vec{v}\|$ is:

$$\|\vec{v}\| = \sqrt{3^2 + 4^2} = \sqrt{9 + 16} = \sqrt{25} = 5$$

Step 3: Verify the Cauchy-Schwarz inequality

The inequality states:

$$|\vec{u} \cdot \vec{v}| \leq \|\vec{u}\| \; \|\vec{v}\|$$

Substitute the values:

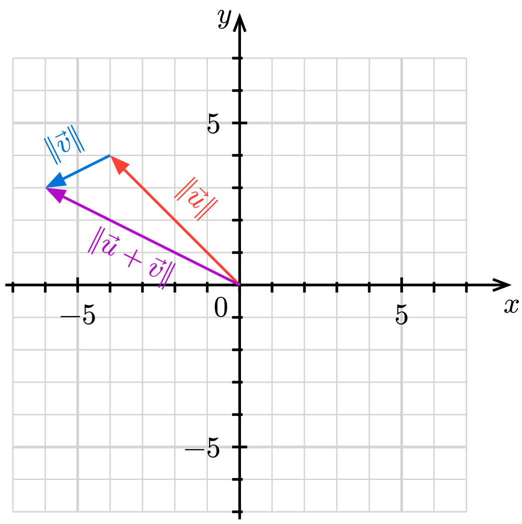
$$|11| \leq 5\sqrt{5} = 11.18$$

Since $11 \leq 11.18$, the inequality holds.

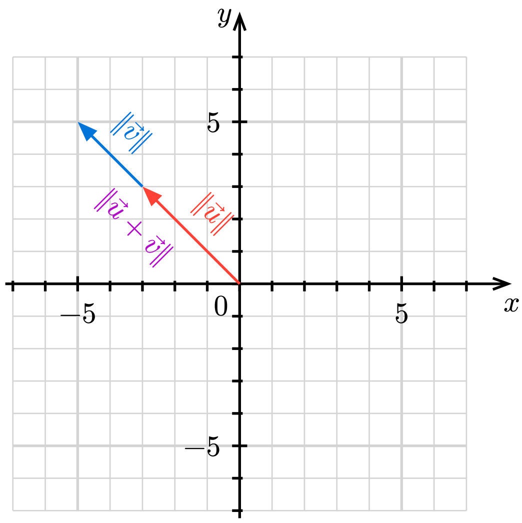
The Cauchy-Schwarz inequality is satisfied for the vectors $\vec{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

11.1.4. Vector Triangle Inequality

$$\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$$



$$\|\vec{u} + \vec{v}\| = \|\vec{u}\| + \|\vec{v}\|$$

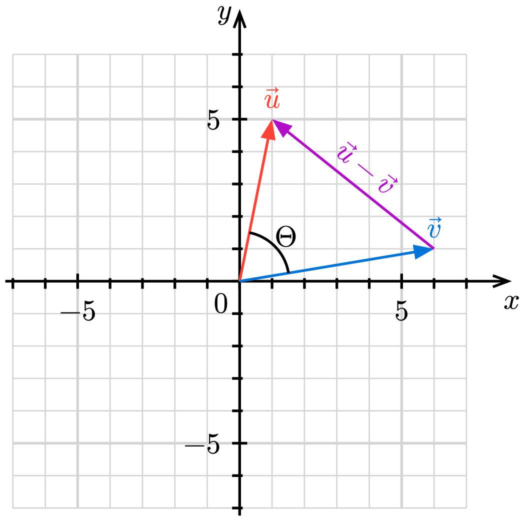


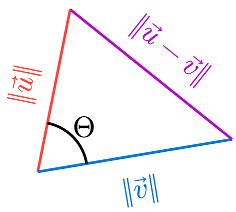
$$\vec{u} = c\vec{v} \qquad c > 0$$

11.2. Angles Between Vectors

The scalar $\|\vec{u}\|$ is the length of the vector $\vec{u}$

Say $\vec{u}, \vec{v} \in \mathbb{R}^n$





Law of Cosines

$$c^2 = a^2 + b^2 - 2ab \cdot \cos(C)$$

Where:

- a, b and c: lengths of the sides of a triangle
- C: angle opposite the side c

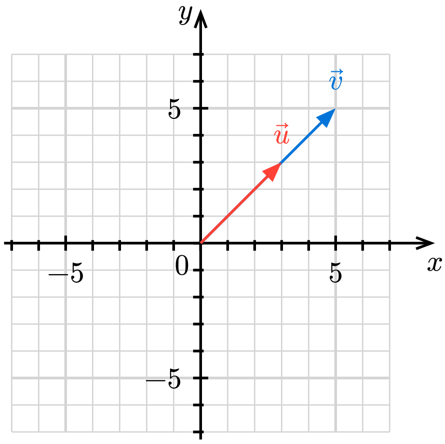
$$\begin{aligned}
\|\vec{u}-\vec{v}\|^2 &= \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\ \|\vec{u}\|\ \|\vec{v}\| \cdot \cos(\Theta) \\
(\vec{u}-\vec{v}) \cdot (\vec{u}-\vec{v}) &= \\
\vec{u}\cdot\vec{u} - \vec{u}\cdot\vec{v} - \vec{v}\cdot\vec{u} + \vec{v}\cdot\vec{v} &= \\
\cancel{\|\vec{u}\|^2}\cancel{-2}(\vec{u}\cdot\vec{v}) + \cancel{\|\vec{v}\|^2} &= \cancel{\|\vec{u}\|^2} + \cancel{\|\vec{v}\|^2}\cancel{-2}\ \|\vec{u}\|\ \|\vec{v}\| \cdot \cos(\Theta)
\end{aligned}$$

$$\boxed{\vec{u}\cdot\vec{v} = \|\vec{u}\|\ \|\vec{v}\| \cdot \cos(\Theta)}$$

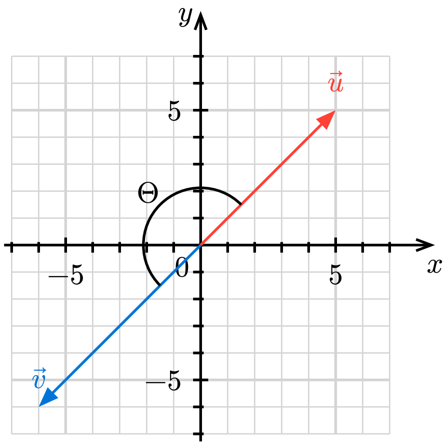
$$\boxed{\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\ \|\vec{v}\|} = \cos(\Theta)}$$

$$\boxed{\Theta = \arccos\left(\frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\ \|\vec{b}\|}\right)}$$

So, if $\vec{u}$ is a scalar multiple of $\vec{v}$ ($\vec{u} = c\vec{v}$) where $c > 0$, then $\Theta = 0°$



And, if $\vec{u}$ is a scalar multiple of $\vec{v}$ ($\vec{u} = c\vec{v}$) where $c < 0$, then $\Theta = 180°$



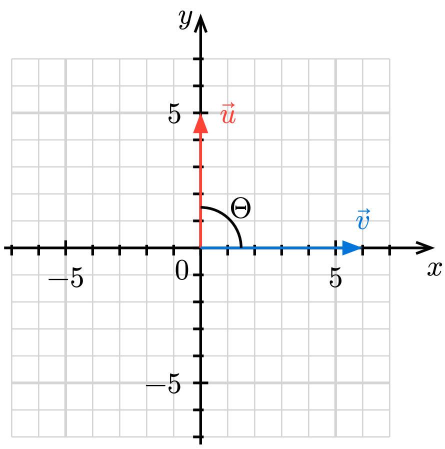
$\vec{u}$ and $\vec{v}$ are perpendicular if the angle Θ between them is $90°$

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \, \|\vec{v}\| \cdot \cos(90)$$
$$\vec{u} \cdot \vec{v} = 0$$

If $\vec{u}$ and $\vec{v}$ are perpendicular, then $\vec{u} \cdot \vec{v} = 0$

If $\vec{u}$ and $\vec{v}$ are non-zer0 and $\vec{u} \cdot \vec{v} = 0$, then they are perpendicular

If $\vec{u} \cdot \vec{v} = 0$ then $\vec{u}$ and $\vec{v}$ are **orthogonal**.



11.2.1. Plane in $\mathbb{R}^3$

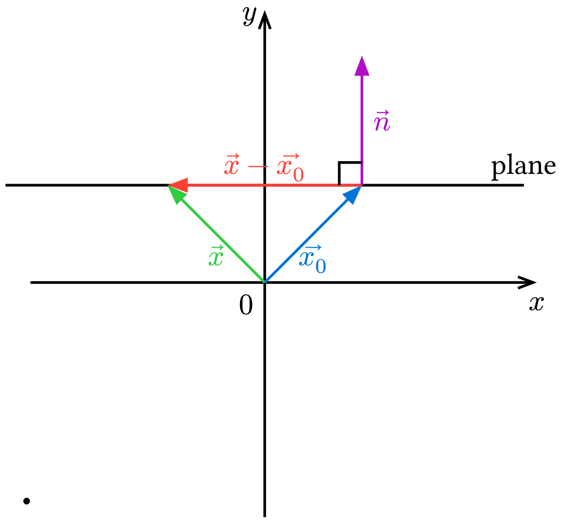
Plane: Each point (x, y, z) on the satifies the equation

$$ax + by + cy = d$$

Normal Vector: vector that is perpendicular (orthogonal) to a plane, line, or curve, at a specific point

1. **Plane**

If a plane is defined by the equation $ax + by + cz = d$, the vector $\vec{n} = \langle a, b, c \rangle$ is a normal vector to the plane because it is perpendicular to any vector that lies in the plane



-

$$\vec{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad \vec{x_0} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \qquad \vec{n} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}$$

$$\vec{x} - \vec{x_0} = \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix}$$

$\vec{x} - \vec{x_0}$ is a vector that lies on the plane, then $\vec{n}$ is normal if:

$$\vec{n} \cdot \vec{x} - \vec{x_0} = 0$$

$$\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \cdot \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} = 0$$

$$\boxed{n_1(x - x_0) + n_2(y - y_0) + n_3(z - z_0) = 0}$$

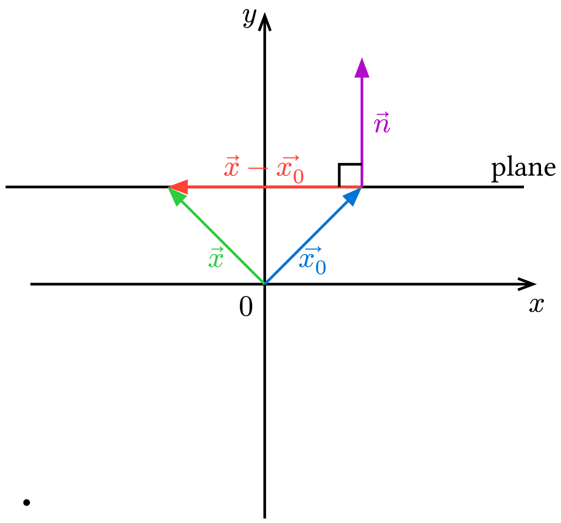
Find the equation of the plane given the point on the plane $\vec{x_0}$ and a the normal vector $\vec{n}$

$$\vec{n} = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} \qquad \vec{x_0} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$\vec{x} - \vec{x_0} = \begin{bmatrix} x - 1 \\ y - 2 \\ z - 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} x - 1 \\ y - 2 \\ z - 3 \end{bmatrix} = 0$$

$$(x - 1) + 3(y - 2) - 2(z - 3) = 0$$
$$x - 1 + 3y \cancel{- 6} - 2z \cancel{+ 6} = 0$$
$$x + 3y - 2z = 1$$



The normal vector to a plane can be directly obtained from the coefficients of x, y, and z in the plane equation of the form:

$$Ax + By + Cz = D$$

$$\vec{n} = \begin{bmatrix} A \\ B \\ C \end{bmatrix}$$

Find the equation of the normal vector $\vec{n}$ given the equation for the plane:

$$-3x + \sqrt{2}y + 7z = \pi$$
$$\vec{n} = -3\hat{i} + \sqrt{2}\hat{j} + 7\hat{k}$$
$$\vec{n} = \begin{bmatrix} -3 \\ \sqrt{2} \\ 7 \end{bmatrix}$$

2. **Curve**

For a curve described by a function $y = f(x)$, the normal vector at a point on the curve is perpendicular to the tangent line at that point. If the tangent vector has slope $f'(x)$, the normal vector will have a slope of $-\frac{1}{f'(x)}$

11.2.2. Point Distance to Plane

$$\boxed{d = \frac{Ax_0 + By_0 + Cz_0 - D}{\sqrt{A^2 + B^2 + C^2}}}$$

Given a the equation of the plane:

$$1x - 2y + 3z = 5$$

And a point **not** on the plane:

$$(2, 3, 1)$$

Find the shortest path (normal vector) from the plane to the point

$$\begin{aligned} d &= \frac{Ax_0 + By_0 + Cz_0 - D}{\sqrt{A^2 + B^2 + C^2}} \\ &= \frac{1 \cdot 2 - 2 \cdot 3 + 3 \cdot 1 - 5}{\sqrt{1^2 + 2^2 + 3^2}} \\ &= \frac{2 - 6 + 3 - 5}{\sqrt{1 + 4 + 9}} \\ &= -\frac{6}{\sqrt{14}} \end{aligned}$$

11.2.3. Distance Between Planes

11.2.4. Cross Product

Only defined in $\mathbb{R}^3$

Returns a vector orthogonal to the two vectors

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$\vec{c} = \vec{a} \times \vec{b}$$

$$\vec{c} = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix}$$

$$\vec{a} = \begin{bmatrix} 1 \\ -7 \\ 1 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} 5 \\ 2 \\ 4 \end{bmatrix}$$

$$\vec{c} = \vec{a} \times \vec{b}$$

$$\vec{c} = \begin{bmatrix} -7 \cdot 4 - 1 \cdot 2 \\ 1 \cdot 5 - 1 \cdot 4 \\ 1 \cdot 2 - (-7) \cdot 5 \end{bmatrix} = \begin{bmatrix} -30 \\ 1 \\ 37 \end{bmatrix}$$

$\vec{c}$ is orthogonal to both $\vec{a}$ and $\vec{b}$

Proof: $\vec{c}$ is orthogonal to $\vec{a}$ and $\vec{b}$

When the dot product of two vectors is equal to 0, it means that the two vectors are perpendicular (or orthogonal) to each other

1. Orthogonal to Vector $\vec{a}$

$$\vec{c} = \begin{bmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

$$a_1a_2b_3 - a_1a_3b_2 + a_2a_3b_1 - a_2a_1b_3 + a_3a_1b_2 - a_3a_2b_1$$

$$\cancel{a_1a_2b_3} - \cancel{a_1a_3b_2} + \cancel{a_2a_3b_1} - \cancel{a_2a_1b_3} + \cancel{a_3a_1b_2} - \cancel{a_3a_2b_1} = 0$$

2. Orthogonal to Vector $\vec{b}$

$$\vec{c} = \begin{bmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$b_1a_2b_3 - b_1a_3b_2 + b_2a_3b_1 - b_2a_1b_3 + b_3a_1b_2 - b_3a_2b_1$$

$$\cancel{b_1a_2b_3} - \cancel{b_1a_3b_2} + \cancel{b_2a_3b_1} - \cancel{b_2a_1b_3} + \cancel{b_3a_1b_2} - \cancel{b_3a_2b_1} = 0$$

11.2.5. Proof: Relationship Between Cross Product and Sin of Angle

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$\vec{c} = \vec{a} \times \vec{b}$$

$$\vec{c} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \times \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{bmatrix}$$

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \ \|\vec{b}\| \cos(\Theta)$$

$$\|\vec{a} \times \vec{b}\| = \|\vec{a}\| \ \|\vec{b}\| \sin(\Theta)$$

11.2.6. Dot and Cross Products

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \ \|\vec{b}\| \cos(\Theta)$$

$$\frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \ \|\vec{b}\|} = \cos(\Theta)$$

$$\Theta = \arccos()$$

11.3. Row Echelon Form (REF)

Visual Structure:

1. Pivot (leading 1): The leading entry of each non-zero row is 1

2. Zeros below pivots: Every pivot has zeros below it in its column
3. Rightward movement of pivots: Each leading 1 in a lower row is further to the right than in the row above it
4. Rows of all zeros (if any) are at the bottom of the matrix

$$\left[\begin{array}{cccc|c} 1 & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & 1 & a_{23} & a_{24} & b_2 \\ 0 & 0 & 1 & a_{34} & b_3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Elementary row operations:
1. **Row Swapping**: Swap two rows
2. **Row Multiplication**: Multiply a row by a non-zero scalar
3. **Row Addition / Subtraction**: Add or subtract a multiple of one row from another row

Consider the system of linear equations:

$$\begin{aligned} 2x + y + z &= 8 \\ -3x - y + 2z &= -11 \\ -2x + 1y + 2z &= -3 \end{aligned}$$

The augmented matrix for this system is:

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right]$$

Step 1: Make the leading entry of the first row a 1

We divide the first row by 2 (row multiplication):

$$R_1 \to \frac{1}{2} R_1 = \left[\begin{array}{ccc|c} 1 & 0.5 & 0.5 & 4 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right]$$

Step 2: Eliminate the entries below the first pivot We now want the entries below the first pivot (1 in the first column) to become zeros. We use row addition:

1. $R_2 \to R_2 + 3R_1$
2. $R_3 \to R_3 + 2R_1$

This gives:

$$\left[\begin{array}{ccc|c} 1 & 0.5 & 0.5 & 4 \\ 0 & 0.5 & 3.5 & 1 \\ 0 & 2 & 3 & 5 \end{array}\right]$$

Step 3: Make the leading entry of the second row a 1

We divide the second row by 0.5 (row multiplication):

$$R_2 \to \frac{1}{0.5} R_2 = \left[\begin{array}{ccc|c} 1 & 0.5 & 0.5 & 4 \\ 0 & 1 & 7 & 2 \\ 0 & 2 & 3 & 5 \end{array}\right]$$

Step 4: Eliminate the entry below the second pivot

We now want the entry below the second pivot (1 in the second column) to become zero. We use row addition:

$$R_3 \to R_2 - 2R_2 = \left[\begin{array}{ccc|c} 1 & 0.5 & 0.5 & 4 \\ 0 & 1 & 7 & 2 \\ 0 & 0 & -11 & 1 \end{array}\right]$$

Step 5: Make the leading entry of the third row a 1

We divide the third row by −11 (row multiplication):

$$R_3 \to -\frac{1}{11}R_3 = \left[\begin{array}{ccc|c} 1 & 0.5 & 0.5 & 4 \\ 0 & 1 & 7 & 2 \\ 0 & 0 & 1 & -\frac{1}{11} \end{array}\right]$$

Step 6: Back-substitute to solve for the variables

From the third row:

$$z = -\frac{1}{11}$$

Substitute z into the second row:

$$\begin{gathered} y + 7z = 2 \\ y + 7\left(-\frac{1}{11}\right) = 2 \\ y = 2 + \frac{7}{11} \\ y = \frac{29}{11} \end{gathered}$$

Substitute y and z into the first row:

$$\begin{gathered} x + 0.5y + 0.5z = 4 \\ x + 0.5\left(\frac{29}{11}\right) + 0.5\left(-\frac{1}{11}\right) = 4 \\ x = 4 - \frac{14}{11} + \frac{1}{11} \\ x = \frac{31}{11} \end{gathered}$$

Final Solution:

$$x = \frac{31}{11} \qquad y = \frac{29}{11} \qquad z = -\frac{1}{11}$$

11.3.1. Solution Types in Linear Systems: Unique, Infinite, or None

1. Unique Solution

$$\left[\begin{array}{cccc|c} 1 & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & 1 & a_{23} & a_{24} & b_2 \\ 0 & 0 & 1 & a_{34} & b_3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

2. No Solution

$$0 = a$$

3. No Unique Solution (Infinite Number of Solutions)

Column 2 and 4 indicate free variables x_2 and x_4 becuase they have no pivot entries

$$\left[\begin{array}{cccc|c} 1 & a_{12} & a_{13} & a_{14} & b_1 \\ 0 & 0 & 1 & a_{24} & b_2 \\ 0 & 0 & 1 & a_{34} & b_3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

11.3.2. Special Cases

Rows of all zeros appear in row echelon form (REF) in the following situations:

1. Dependent Equations

Some equations are multiples or linear combinations of others

$$2x + 4y = 8$$
$$x + 2y = 4$$

Augmented matrix:

$$\left[\begin{array}{cc|c} 2 & 4 & 8 \\ 1 & 2 & 4 \end{array}\right]$$

After row reduction:

$$\left[\begin{array}{cc|c} 1 & 2 & 4 \\ 0 & 0 & 0 \end{array}\right]$$

2. Underdetermined Systems

Number of variables is greater than the number of independent equations

$$x + y + z = 2$$
$$2x + 3y + z = 5$$

Augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 2 & 3 & 1 & 5 \end{array}\right]$$

After row reduction:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

3. Inconsistent Systems

No solution exists

Rows of zeros on the left side (coefficients of the variables) and a non-zero entry on the right side (augmented column)

$$x + y = 3$$
$$2x + 2y = 7$$

Augmented matrix:

$$\left[\begin{array}{cc|c} 1 & 1 & 3 \\ 2 & 2 & 7 \end{array}\right]$$

After row reduction:

$$\left[\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & 0 & 1 \end{array}\right]$$

12. Matrices

$m \times n$ matrix $\boldsymbol{A}$

- m: rows
- n: columns

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

12.1. Matrix-Vector Products

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\boldsymbol{A}\vec{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

For the dot product to be defined, the number of columns in the matrix $\boldsymbol{A}$ (which is n) must match the number of elements in the vector $\vec{x}$ (also n).

The result of multiplying matrix $\boldsymbol{A}$ and vector $\vec{x}$ will be a column vector with dimensions $m \times 1$, where m is the number of rows in the matrix $\boldsymbol{A}$

$$(m \times n) \cdot (n \times 1) = m \times 1$$

1. As Row vectors

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\vec{a}^T = [a_1, a_2, ..., a_n]$$

$$\vec{b}^T = [b_1, b_2, ..., b_n]$$

$$\boldsymbol{A} = \begin{bmatrix} [a_1, a_2, ..., a_n] \\ [b_1, b_2, ..., b_n] \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} \vec{a} \\ \vec{b} \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\begin{bmatrix} \vec{a}^T \\ \vec{b}^T \end{bmatrix} \cdot \vec{x} = \begin{bmatrix} \vec{a} \cdot \vec{x} \\ \vec{b} \cdot \vec{x} \end{bmatrix}$$

2. As Column Vectors

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} & \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} \vec{a} & \vec{b} \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\boldsymbol{A}\vec{x} = x_1\vec{a} + x_2\vec{b}$$

12.2. Null Space

The null space (or kernel) of a matrix $\boldsymbol{A}$ is the set of all vectors $\boldsymbol{x}$ that satisfy the equation:

$$\boldsymbol{A}\vec{x} = \mathbf{0}$$

Where:

- $\boldsymbol{A}$: $m \times n$ matrix
- $\vec{x}$: n-dimansional vector
- $\mathbf{0}$: zero vecor in $\mathbb{R}^m$

$$N(\boldsymbol{A}) = N(\text{rref}(\boldsymbol{A})) = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$

We want to find the null space of A, which consists of all vectors $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ that satisfy:

$$\boldsymbol{A}\vec{x} = \mathbf{0}$$

This expands to the following system of linear equations:

$$\begin{cases} 1x_1 + 1x_2 + 1x_3 + 1x_4 = 0 \\ 1x_1 + 2x_2 + 3x_3 + 4x_4 = 0 \\ 4x_1 + 3x_2 + 2x_3 + 1x_4 = 0 \end{cases}$$

This can be represented as the augmented matrix:

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 0 \\ 4 & 3 & 2 & 1 & 0 \end{array}\right]$$

12.2.1. Column Space

The **columns space** (or range) of matrix A is span of its columns vectors

If the matrix A has columns $\vec{a}_1, \vec{a}_2, ..., \vec{a}_n$, then the column space of A is defined as:

$$\text{Col}(A) = \{\vec{y} \in \mathbb{R}^m \mid \vec{y} = A\vec{x} \text{ for some } \vec{x} \in \mathbb{R}^n\}$$

or equivalently,

$$\text{Col}(A) = \text{span}(\{\vec{a}_1, \vec{a}_2, ..., \vec{a}_n\})$$

Consider the simple example of a 2×2 matrix:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$$

The matrix has two columns:

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{a}_2 = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$$

The column space, denoted $\text{Col}(A)$, is the span of these two vectors:

$$\text{Col}(A) = \text{span}\left(\left\{\begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}2\\6\end{bmatrix}\right\}\right)$$

Finding the Column Space

We observe that the two columns $\vec{a}_1$ and $\vec{a}_2$ are **linearly dependent**:

$$\vec{a}_2 = k\vec{a}_1$$

This means that $\vec{a}_2$ is a scalar multiple of $\vec{a}_1$, the the two columns are **linearly dependent**. As a result, the column space is spanned by just one vector, $\vec{a}_1$, because any linear combination of $\vec{a}_1$ and $\vec{a}_2$ can be reduced to a multiple of $\vec{a}_1$.

Therefore, the column space of A is:

$$\text{Col}(A) = \text{span}\left(\left\{\begin{bmatrix}1\\3\end{bmatrix}\right\}\right)$$

which represents all vectors of the form:

$$c\begin{bmatrix}1\\3\end{bmatrix} = \begin{bmatrix}c\\3c\end{bmatrix} \qquad \text{for any scalar } c$$

In other words, the column space is a line in $\mathbb{R}^2$ through the origin in the direction of

Rank of A

The rank of A, which is the **dimension of its column space**, is 1 because there is only one linearly independent column

This means the column space is the span of the columns of A, or all vectors that can be formed by taking linear combinations of the columns of A.

12.2.2. Dimension of a Subspace

Number of elements in a basis for the subspace

12.2.3. Nullity

Dimension of the **Null Space**

$$\dim(N(\boldsymbol{A}))$$

The nullity of A: number of non-pivot columns (i.e., free variables) in the rref of A

12.2.4. Rank

Dimension of the column **space**

$$\text{rank}(A) = \dim(C(A))$$

12.2.5. Matrix Representation of Systems of Equations

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m = b_n \end{gathered}$$

Coefficient Matrix (A):

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

Variable Vector (x):

$$\mathrm{x} \ = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

Constant Vector (b):

$$\mathrm{b} \ = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\mathrm{Ax} = b$$

The system of equations:

$$\begin{aligned} 2x_1 + 3x_2 + 5x_3 &= 100 \\ 4x_1 + 2x_2 + 1x_3 &= 80 \\ 1x_1 + 5x_2 + 2x_3 &= 60 \end{aligned}$$

Can be represented as a matrix equation:

$$\begin{bmatrix} 2 & 3 & 5 \\ 4 & 2 & 1 \\ 1 & 5 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 100 \\ 80 \\ 60 \end{bmatrix}$$

12.3. Matrix Multiplication

$m \times n$ matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

$n \times p$ matrix:

$$B = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix}$$

Compute Each Element of Result Matrix C

$$c_{ij} = \sum_{k=1}^{n} = a_{ik} b_{kj}$$

Let A be an $n \times m$ matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Let B an $p \times n$ matrix:

$$B = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}$$

Calculate Each Element of C

$$\begin{aligned} c_{11} &= (1 \cdot 7) + (2 \cdot 9) + (3 \cdot 11) = 58 \\ c_{12} &= (1 \cdot 8) + (2 \cdot 10) + (3 \cdot 12) = 64 \\ c_{21} &= (4 \cdot 7) + (5 \cdot 9) + (6 \cdot 11) = 138 \\ c_{22} &= (4 \cdot 8) + (5 \cdot 10) + (6 \cdot 12) = 154 \end{aligned}$$

C is a $m \times p$ matrix

$$C = \begin{bmatrix} 58 & 64 \\ 139 & 154 \end{bmatrix}$$

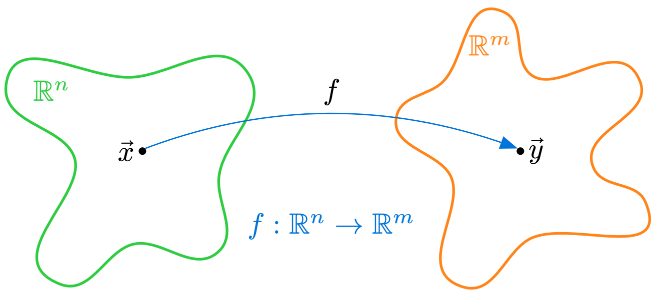
12.4. Linear Transformation

12.4.1. Functions

A function f that maps elements from a set X (the domain) to a set Y (the codomain):

$$f : X \to Y$$

- Domain: The set X contains all possible inputs for the function f
- Codomain: The set Y is the space where all possible outputs of f reside, though not every element in Y must be an output of f



If

$$f : \mathbb{R} \to \mathbb{R}$$

is defined by

$$f(x) = x^2$$
$$f : x \mapsto x^2$$

then:

- **Domain**: $X = \mathbb{R}$, any real number $((\infty, \infty))$
- **Codomain**: $Y = \mathbb{R}$, any real number $((\infty, \infty))$
- **Range**: the subset of the codomain ($\mathbb{R}$), $[0, \infty)$

12.5. Vector Transformation

A function f that maps an n-dimensional vector in $\mathbb{R}^n$ to an m-dimensional vector in $\mathbb{R}^m$:

1. Function Definition

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

This means f takes as input a vector in $\mathbb{R}^n$ (an n-dimensional space of real numbers) and maps it to a vector in $\mathbb{R}^m$ (an m-dimensional space of real numbers).

2. Input Vector $\vec{x}$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{where} \quad x_1, x_2, ..., x_n \in \mathbb{R}$$

Here, $\vec{x}$ is an n-dimensional vector, and each component x_i is a real number

3. Input Vector $\vec{y}$

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \quad \text{where} \quad y_1, y_2, ..., y_m \in \mathbb{R}$$

The output $\vec{y}$ is an m-dimensional vector, and each component y_i is also a real number

Summary

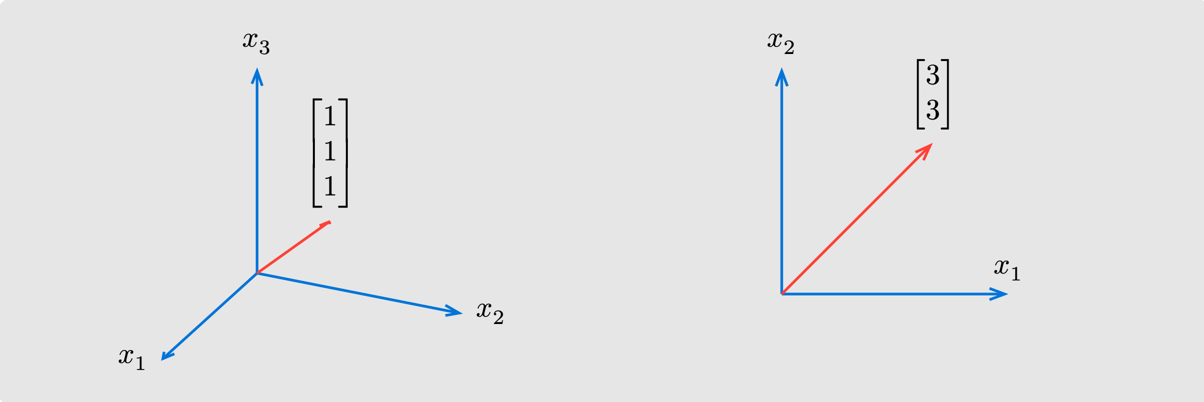
The function f takes an n-dimensional vector of real numbers as input and produces an m-dimensional vector of real numbers as output

$$f(x_1, x_2, x_3) = (x_1 + 2x_2, 3x_3)$$

$$f : \mathbb{R}^3 \to \mathbb{R}^2$$

$$f\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ 3x_3 \end{bmatrix}$$

$$f\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$



12.6. Linear Transformation

$$T : \mathbb{R}^n \to \mathbb{R}^m$$

$$\vec{a}, \vec{b} \in \mathbb{R}^n$$

For a transformation **linear** it must satisfy two conditions:

1. Additivity (or linearity of addition)

$$\boxed{T\left(\vec{a} + \vec{b}\right) = T(\vec{a}) + T\left(\vec{b}\right)}$$

Let's consider a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$

$$T\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \begin{bmatrix}2x\\3y\end{bmatrix}$$

Now let's take two vectors in $\mathbb{R}^2$:

$$\vec{a} = \begin{bmatrix}1\\2\end{bmatrix} \qquad \vec{b} = \begin{bmatrix}3\\1\end{bmatrix}$$

Then the additivity property can be verified as follows:

1. First, find $T(\vec{a}) + T\left(\vec{b}\right)$ separately:

$$T(\vec{a}) = T\left(\begin{bmatrix}1\\2\end{bmatrix}\right) = \begin{bmatrix}2 \cdot 1\\3 \cdot 2\end{bmatrix} = \begin{bmatrix}2\\6\end{bmatrix}$$

$$T\left(\vec{b}\right) = T\left(\begin{bmatrix}3\\1\end{bmatrix}\right) = \begin{bmatrix}2 \cdot 3\\3 \cdot 1\end{bmatrix} = \begin{bmatrix}6\\3\end{bmatrix}$$

$$T(\vec{a}) + T\left(\vec{b}\right) = \begin{bmatrix}2\\6\end{bmatrix} + \begin{bmatrix}6\\3\end{bmatrix} = \begin{bmatrix}2+6\\6+3\end{bmatrix} = \begin{bmatrix}8\\9\end{bmatrix}$$

2. Next, find $T\left(\vec{a} + \vec{b}\right)$:

$$\vec{a} + \vec{b} = \begin{bmatrix}1\\2\end{bmatrix} + \begin{bmatrix}3\\1\end{bmatrix} = \begin{bmatrix}4\\3\end{bmatrix}$$

$$T\left(\vec{a} + \vec{b}\right) = T\left(\begin{bmatrix}4\\3\end{bmatrix}\right) = \begin{bmatrix}2 \cdot 4\\3 \cdot 3\end{bmatrix} = \begin{bmatrix}8\\9\end{bmatrix}$$

Since $T\left(\vec{a}+\vec{b}\right)=T(\vec{a})+T\left(\vec{b}\right)$, this confirms the additivity (linearity of addition) property of the transformation T

2. Homogeneity (or linearity of scalar multiplication):

$$\boxed{T(c\vec{a})=cT(\vec{a})}$$

Let's consider a linear transformation $T:\mathbb{R}^2\to\mathbb{R}^2$

$$T\left(\begin{bmatrix}x\\y\end{bmatrix}\right)=\begin{bmatrix}2x\\3y\end{bmatrix}$$

Now let's take one vectors in $\mathbb{R}^2$:

$$\vec{a}=\begin{bmatrix}1\\2\end{bmatrix}$$

and a scalar $c=3$

Then the homogeneity property can be verified as follows:

1. First, find $cT(\vec{a})$

$$T(\vec{a})=T\left(\begin{bmatrix}1\\2\end{bmatrix}\right)=\begin{bmatrix}2\cdot 1\\3\cdot 2\end{bmatrix}=\begin{bmatrix}2\\6\end{bmatrix}$$

$$cT(\vec{a})=3\cdot\begin{bmatrix}2\\6\end{bmatrix}=\begin{bmatrix}3\cdot 2\\3\cdot 6\end{bmatrix}=\begin{bmatrix}6\\18\end{bmatrix}$$

2. Then, find $T(c\vec{a})$ by c to get $c\vec{a}$

$$c\vec{a}=3\cdot\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}3\cdot 1\\3\cdot 2\end{bmatrix}=\begin{bmatrix}3\\6\end{bmatrix}$$

$$T(c\vec{a})=T\left(\begin{bmatrix}3\\6\end{bmatrix}\right)=\begin{bmatrix}2\cdot 3\\3\cdot 6\end{bmatrix}=\begin{bmatrix}6\\18\end{bmatrix}$$

Since $T(c\vec{a})=cT(\vec{a})$, this confirms the homogeneity (linearity of scalar multiplication) property of the transformation T

12.7. Matrix Vector Products

Matrix product with vector is always a linear transformation

$$T:\mathbb{R}^n\to\mathbb{R}^m$$
$$T(\vec{x})=A\vec{x}$$

$$A = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$A\vec{x} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 v_1 + x_2 v_2 + \dots + x_n v_n$$

1. Additivity (or linearity of addition)

$$\boxed{T\left(\vec{a} + \vec{b}\right) = T(\vec{a}) + T\left(\vec{b}\right)}$$

$$\begin{aligned} A \cdot \left(\vec{a} + \vec{b}\right) = A \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix} &= (a_1 + b_1)v_1 + (a_2 + b_2)v_2 + \dots + (a_n + b_n)v_n \\ &= a_1 v_1 + b_1 v_1 + a_2 v_2 + b_2 v_2 + \dots + a_n v_n + b_n v_n \\ &= (a_1 v_1 + a_2 v_2 + \dots + a_n v_n) + (b_1 v_1 + b_2 v_2 + \dots + b_n v_n) \\ &= A \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + A \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \end{aligned}$$

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \qquad \boldsymbol{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \boldsymbol{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

1. Calculate $A(\boldsymbol{u} + \boldsymbol{v})$

$$\begin{aligned} \boldsymbol{u} + \boldsymbol{v} &= \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 4 \\ 6 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} A(\boldsymbol{u} + \boldsymbol{v}) &= \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix} \\ &= \begin{bmatrix} (2 \cdot 1) + (1 \cdot 6) \\ (0 \cdot 4) + (3 \cdot 6) \end{bmatrix} \\ &= \boxed{\begin{bmatrix} 14 \\ 18 \end{bmatrix}} \end{aligned}$$

2. Calculate $A\boldsymbol{u} + A\boldsymbol{v}$

$$
\begin{aligned}
\boldsymbol{Au} &= \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\
&= \begin{bmatrix} (2 \cdot 1) + (1 \cdot 2) \\ (0 \cdot 1) + (3 \cdot 2) \end{bmatrix} \\
&= \begin{bmatrix} 4 \\ 6 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{Av} &= \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \\
&= \begin{bmatrix} (2 \cdot 3) + (1 \cdot 4) \\ (0 \cdot 3) + (3 \cdot 4) \end{bmatrix} \\
&= \begin{bmatrix} 10 \\ 12 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{Au} + \boldsymbol{Av} &= \begin{bmatrix} 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 10 \\ 12 \end{bmatrix} \\
&= \boxed{\begin{bmatrix} 14 \\ 18 \end{bmatrix}}
\end{aligned}
$$

2. Homogeneity (or linearity of scalar multiplication):

$$
\boxed{T(c\vec{a}) = cT(\vec{a})}
$$

$$
\begin{aligned}
A \cdot (c\vec{a}) &= \begin{bmatrix} \\ v_1 & v_2 & \dots & v_n \\ \\ \end{bmatrix} \begin{bmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{bmatrix} \\
&= ca_1 v_1 + ca_2 v_2 + \dots + ca_n v_n \\
&= c\underbrace{(a_1 v_1 + a_2 v_2 + \dots + a_n v_n)}_{A\vec{a}}
\end{aligned}
$$

$$
A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \qquad \boldsymbol{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad c = 5
$$

1. Calculate $A(c\boldsymbol{v})$

$$
c v = 5 \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 15 \\ 20 \end{bmatrix}
$$

$$
\begin{aligned}
A(c\boldsymbol{v}) &= \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 15 \\ 20 \end{bmatrix} \\
&= \begin{bmatrix} (2 \cdot 15) + (1 \cdot 20) \\ (0 \cdot 15) + (3 \cdot 20) \end{bmatrix} \\
&= \boxed{\begin{bmatrix} 50 \\ 60 \end{bmatrix}}
\end{aligned}
$$

2. Calculate $c(A\boldsymbol{v})$

$$A\boldsymbol{v} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$
$$= \begin{bmatrix} (2 \cdot 3) + (1 \cdot 4) \\ (0 \cdot 3) + (3 \cdot 4) \end{bmatrix}$$
$$= \begin{bmatrix} 10 \\ 12 \end{bmatrix}$$

$$c(A\boldsymbol{v}) = 5 \cdot \begin{bmatrix} 10 \\ 12 \end{bmatrix}$$
$$= \boxed{\begin{bmatrix} 50 \\ 60 \end{bmatrix}}$$

12.8. Linear transformations as matrix vector products

The $n \times n$ matrix I_n:

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$I_n\vec{x} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Standard Basis

$$I_n\vec{x} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \underbrace{0}_{e_1} & \underbrace{0}_{e_2} & \underbrace{0}_{e_3} & \dots & \underbrace{1}_{e_n} \end{bmatrix}$$

$\{e_1, e_2, \dots, e_n\}$ is the standard basis for $\mathbb{R}^n$

$$I_n \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a_1 \vec{e}_1 + a_2 \vec{e}_2 + ... + a_n \vec{e}_n$$

$$= a_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + ... + a_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ a_2 \\ \vdots \\ 0 \end{bmatrix} + ... + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ a_n \end{bmatrix}$$

12.9. Image of a subset under transformation

$$\vec{x}_0 = \begin{bmatrix} -2 \\ -2 \end{bmatrix} \qquad \vec{x}_1 = \begin{bmatrix} -2 \\ 2 \end{bmatrix} \qquad \vec{x}_2 = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

$$L_0 = \{\vec{x}_0 + t(\vec{x}_1 - \vec{x}_0) \mid 0 \leq t \leq 1\}$$
$$L_1 = \{\vec{x}_1 + t(\vec{x}_2 - \vec{x}_1) \mid 0 \leq t \leq 1\}$$
$$L_2 = \{\vec{x}_2 + t(\vec{x}_0 - \vec{x}_2) \mid 0 \leq t \leq 1\}$$

The triangle T can be defines as the set of these points:

$$S = \{L_0, L_1, L_2\}$$

Let's define a transformation

$$T : \mathbb{R}^2 \to \mathbb{R}^2$$

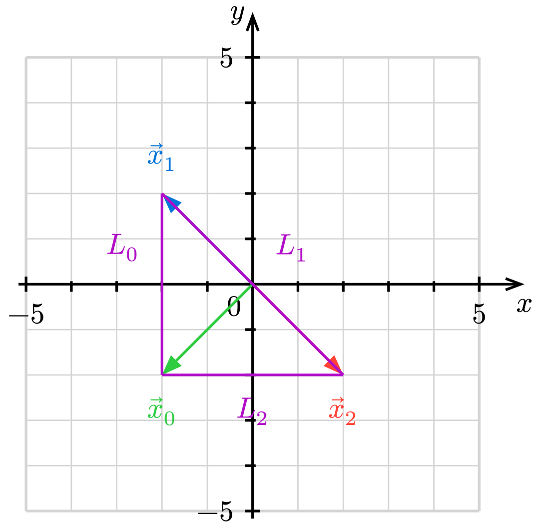
$$T(\vec{x}) = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

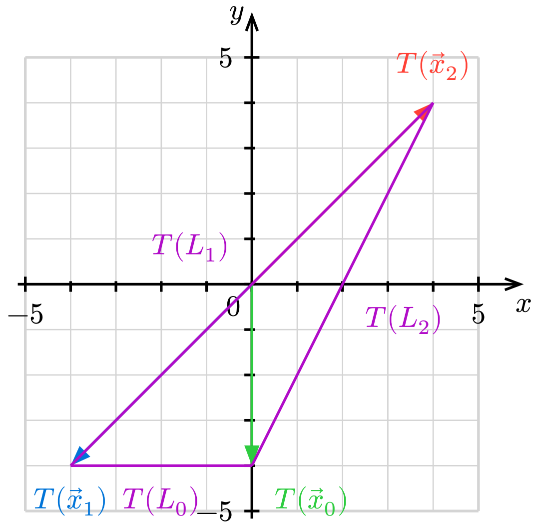
$$\begin{aligned} T(L_0) &= \{T(\vec{x}_0 + t(\vec{x}_1 - \vec{x}_0)) \mid 0 \leq t \leq 1\} \\ &= \{T(\vec{x}_0) + T(t(\vec{x}_1 - \vec{x}_0)) \mid 0 \leq t \leq 1\} \\ &= \{T(\vec{x}_0) + tT(\vec{x}_1 - \vec{x}_0) \mid 0 \leq t \leq 1\} \\ &= \{T(\vec{x}_0) + tT(\vec{x}_1) - T(\vec{x}_0) \mid 0 \leq t \leq 1\} \end{aligned}$$

$$T(\vec{x}_0) = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ -4 \end{bmatrix}$$

$$T(\vec{x}_1) = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} -4 \\ -4 \end{bmatrix}$$

$$T(\vec{x}_2) = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$





$T(L_0)$ is the image of L_0 under T

$T(S)$ is the image of S under T

12.10. Image of a transformation

The **image** of a transformation T is defined as:

$$\text{im}(T) = \{T(\vec{x}) \mid \vec{x} \in \mathbb{R}^n\}$$

or, equivalently,

$$T(\mathbb{R}^n)$$

$T(\mathbb{R}^n)$ is the image of $\mathbb{R}^2$ under T

This is the set of all possible outputs when T is applied to vectors in $\mathbb{R}^n$

Undertanding $T(\mathbb{R}^n)$

1. Whole space transformation

   The image of $\mathbb{R}^n$ under T is the complete set of transformed vectors, often denoted as $\text{im}(T)$

2. Subset Transformation

   For any subset $V \subseteq \mathbb{R}^n$, the image of V under T is the set of transformed vectors from V

Matrix representation of T

If T is represented by a $m \times n$ matrix A, then:

$$T(\vec{x}) = \{A\vec{x} \mid \vec{x} \in \mathbb{R}^n\}$$

where:

- A is the matrix associated with T
- $\vec{x} \in \mathbb{R}^n$ represents a vector in the input space

Transformation in therms of columns of A

If $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \end{bmatrix}$, then for $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$:

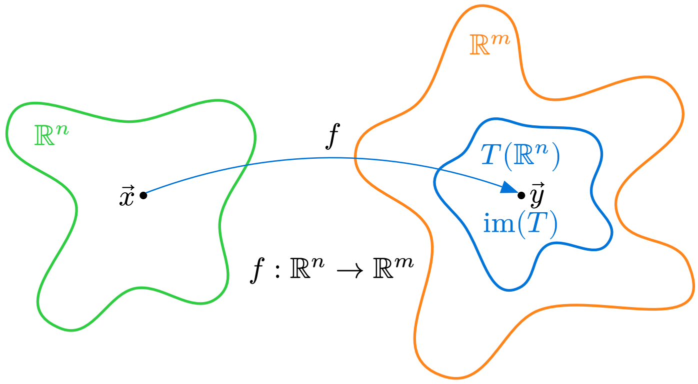
$$A\vec{x} = x_1\vec{a}_1 + x_2\vec{a}_2 + \ldots + x_n\vec{a}_n$$

Column Space of A

The image of T (or $\text{im}(T)$) is the column space of A:

$$C(A) = \text{span}(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n)$$

This is the set of all possible linear combinations of the columns of A, and thus represents all possible outputs of the transformation T



Supose we have a matrix A:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Matrix A defines the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that for any vector $\vec{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right] \in \mathbb{R}^2$, the image under T is:

$$T(\vec{x}) = A\vec{x} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Calculating Transformation

To see what T does to vectors in $\mathbb{R}^2$, let's compute a few specific examples:

1. for $\vec{x} = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = A\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

2. for $\vec{x} = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$

$$T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = A\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Image of T

The image of T, $\text{im}(T)$, is the set of all linear combinations of the vectors $\left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right]$:

$$\text{im}(T) = \text{span}\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}\right)$$

Thus, any vector in the image of T can be written as:

$$y = x_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 3x_2 \end{bmatrix}$$

where $x_1, x_2 \in \mathbb{R}$.

Column space interpretation

The image of T is all the vectors in $\mathbb{R}^2$ that can be formed as linear combinations of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

12.11. Preimage of a set

The preimage of a set S under a function T, denoted $T^{-1}(S)$, is the set of all elements in the domain of T the domain of T that map to elements in S under the transformation T.

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a function from a set $\mathbb{R}^n$ to a set $\mathbb{R}^m$, and $S \subseteq \mathbb{R}^m$ is a subset of a target space, the the preimage of S under T is:

$$T(-1)(S) = \{\vec{x} \in \mathbb{R}^n \mid T(\vec{x}) \in S\}$$

This means that $T(-1)(S)$ consists of all elements in $\mathbb{R}^n$ that, when transformed by T, end up in S.

For any subset $S \subseteq \mathbb{R}^m$, the preimage $T^{-1}(S)$ collects all points in the domain that end up in S after applying T. If S is a single point, the preimage will be the set of all points in the domain that map to that specific point (this could be empty, a single point, or even a set of points, depending on the function).

Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by the matrix:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

This transformation T maps any vector $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ in $\mathbb{R}^2$ to:

$$T(\vec{x}) = A\vec{x} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 3x_2 \end{bmatrix}$$

Now, let's find the preimage of a subset $S \subseteq \mathbb{R}^2$. Suppose we want the preimage of the set $S = \left\{ \begin{bmatrix} 4 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$

The primage of S under T, denoted $T^{-1}(S)$, consists of all vectors $\vec{x} \in \mathbb{R}^2$ such that $T(\vec{x}) \in \left\{ \begin{bmatrix} 4 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$. To find this, we solve for $\vec{x}$ in both cases: $T(\vec{x}) = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$ and $T(\vec{x}) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

Preimage of $\begin{bmatrix} 4 \\ 6 \end{bmatrix}$

$$A\vec{x} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

or

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

Solving each component:

1. $2x_1 = 4 \Rightarrow x_1 = 2$
2. $3x_2 = 6 \Rightarrow x_2 = 2$

Thus, the primage of S is the single point:

$$T^{-1}\left(\begin{bmatrix}4\\6\end{bmatrix}\right) = \left\{\begin{bmatrix}2\\2\end{bmatrix}\right\}$$

Preimage of $\begin{bmatrix}2\\3\end{bmatrix}$

$$A\vec{x} = \begin{bmatrix}2\\3\end{bmatrix}$$

or

$$\begin{bmatrix}2 & 0\\0 & 3\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}2\\3\end{bmatrix}$$

Solving each component:

1. $2x_1 = 2 \Rightarrow x_1 = 1$
2. $3x_2 = 3 \Rightarrow x_2 = 1$

Thus, the primage of S is the single point:

$$T^{-1}\left(\begin{bmatrix}2\\3\end{bmatrix}\right) = \left\{\begin{bmatrix}1\\1\end{bmatrix}\right\}$$

Preimage of the set S

Since $S = \left\{\begin{bmatrix}4\\6\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}\right\}$, the prrimage of S is the union of the preimage of each vector in S:

$$T^{-1}(S) = \left\{\begin{bmatrix}2\\2\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}\right\}$$

12.11.1. Kernel of a Transofrmation

The kernel of a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, denoted $\ker(T)$, is the set of all vectors in $\mathbb{R}^n$ that T maps to the zero vector in $\mathbb{R}^m$. Formally, we define the kernel as:

$$\ker(T) = \left\{\vec{x} \in \mathbb{R}^n \mid T(\vec{x}) = \vec{0}\right\}$$

The consists of all vectors that are "annihilated" by T, resulting in the zero vector after applying T.

The transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by the matrix

$$A = \begin{bmatrix}2 & 0\\0 & 3\end{bmatrix}$$

so that $T(\vec{x}) = A\vec{x} = \begin{bmatrix}2x_1\\3x_2\end{bmatrix}$ for any $\vec{x} = \begin{bmatrix}x_1\\x_2\end{bmatrix} \in \mathbb{R}^2$

To find the kernel of T, we need to find all the vectors $\vec{x} \in \mathbb{R}^2$ that satisfy:

$$T(\vec{x}) = \vec{0} \Rightarrow \begin{bmatrix}2 & 0\\0 & 3\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$

This leads to the system of equations:

1. 3 x_1 = 0 arrow.double x_1 = 0

2. 3 x_2 = 0 arrow.double x_2 = 0

Thus, the only solution is $\vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

12.11.2. Kernel and Null Space

The kernel of T

$$\ker(T) = \mathrm{Null}(A)$$

12.12. Sum and Scalar Multiples of Linear Transformation

12.12.1. Sum

$$T : \mathbb{R}^n \to \mathbb{R}^m \qquad S : \mathbb{R}^n \to \mathbb{R}^m$$

$$(T + S) : \mathbb{R}^n \to \mathbb{R}^m$$

$$A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \end{bmatrix} \quad B = \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \dots & \vec{b}_n \end{bmatrix}$$

$$\begin{aligned} (T + S)(\vec{x}) &= T(\vec{x}) + S(\vec{x}) \\ &= A\vec{x} + B\vec{x} \\ &= x_1\vec{a}_1 + x_2\vec{a}_2 + \dots + x_n\vec{a}_n + x_1\vec{b}_1 + x_2\vec{b}_2 + \dots + x_n\vec{b}_n \\ &= x_1\left(\vec{a}_1 + \vec{b}_1\right) + x_2\left(\vec{a}_2 + \vec{b}_2\right) + \dots + x_n\left(\vec{a}_n + \vec{b}_n\right) \\ &= \begin{bmatrix} \vec{a}_1 + \vec{b}_1 & \vec{a}_2 + \vec{b}_2 & \dots & \vec{a}_n + \vec{b}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \boxed{(A + B)\vec{x}} \end{aligned}$$

12.12.2. Scalar Multiplication

$$T : \mathbb{R}^n \to \mathbb{R}^m$$

$$cT : \mathbb{R}^n \to \mathbb{R}^m$$

$$\begin{aligned} (cT)(\vec{x}) &= c(T(\vec{x})) \\ &= c(x_1\vec{a}_1 + x_2\vec{a}_2 + \dots + x_n\vec{a}_n) \\ &= x_1 c\vec{a}_1 + x_2 c\vec{a}_2 + \dots + x_n c\vec{a}_n \\ &= \begin{bmatrix} c\vec{a}_1 & c\vec{a}_2 & \dots & c\vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \boxed{cA\vec{x}} \end{aligned}$$

12.13. Projection

The projection of a vector $\vec{x}$, onto a line L, denoted as $\text{Proj}_L(\vec{x})$, is a vector that lies on the line L, such that the difference between $\vec{x}$ and its projection, $\text{Proj}_L(\vec{x}) - \vec{x}$, is orthogonal to L

$\text{Proj}_L(\vec{x})$ can be seen as the “shadow”cast by $\vec{x}$ onto L when light shines perpendicularly to L.

$$\text{Proj}_L(\vec{x}) = c\vec{v} = \left(\frac{\vec{x} \cdot \vec{v}}{\vec{v} \cdot \vec{v}}\right)\vec{v}$$

Equivalently

$$\text{Proj}_L(\vec{x}) = c\vec{v} = \left(\frac{\vec{x} \cdot \vec{v}}{\|\vec{v}\|^2}\right)\vec{v}$$

Where:

- $\vec{v}$ is a direction vector for the line L
- $c = \frac{\vec{x} \cdot \vec{v}}{\vec{v} \cdot \vec{v}}$ is a scalar

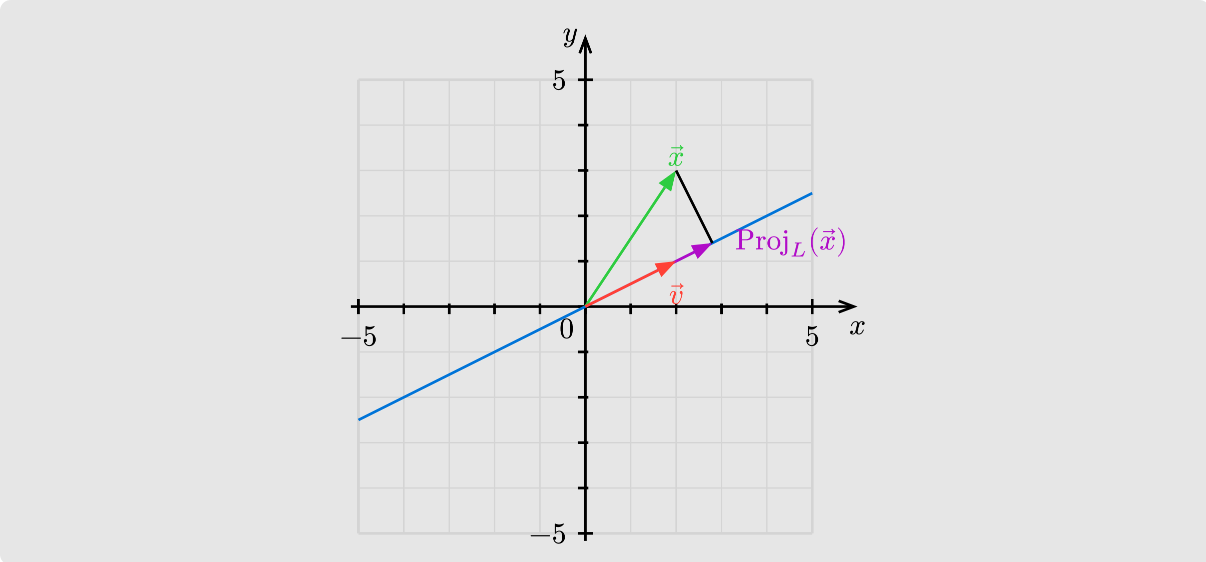
$$\vec{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \qquad \vec{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

1. Define the line L as all vectors of the form $c\vec{v}$, where c is a scalar:

$$\begin{aligned} L &= \{c\vec{v} \mid c \in \mathbb{R}\} \\ &= \left\{c\begin{bmatrix} 2 \\ 1 \end{bmatrix} \mid c \in \mathbb{R}\right\} \end{aligned}$$

2. Compute the projection using

$$\begin{aligned} \text{Proj}_L(\vec{x}) &= \frac{\vec{x} \cdot \vec{v}}{\vec{v} \cdot \vec{v}}\vec{v} \\ &= \frac{\begin{bmatrix} 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix}}{\begin{bmatrix} 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix}}\begin{bmatrix} 2 \\ 1 \end{bmatrix} \\ &= \frac{7}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 2.8 \\ 1.4 \end{bmatrix} \end{aligned}$$



Projection as a Transformation

As a matrix vector product

$$L = \{c\vec{v} \mid c \in \mathbb{R}\}$$

$$\text{Proj}_L : \mathbb{R}^n \to \mathbb{R}^n$$

$$\vec{v} \cdot \vec{v} = \| \vec{v} \|^2$$

If $\vec{v}$ is a unit vector:

$$\|\vec{v}\| = 1$$

Then

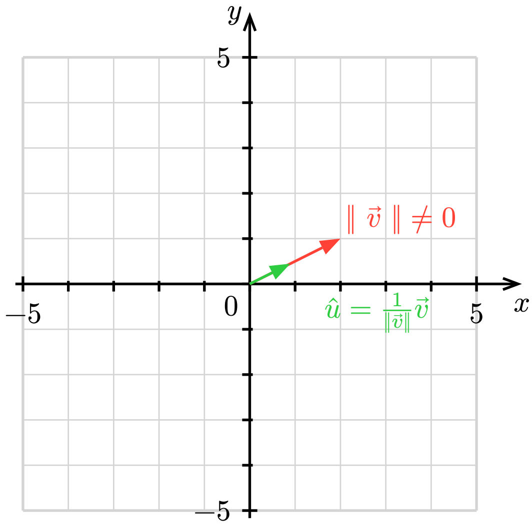
$$\begin{aligned} \text{Proj}_L(\vec{x}) &= \left(\frac{\vec{x} \cdot \vec{v}}{\vec{v} \cdot \vec{v}}\right)\vec{v} \\ &= \left(\frac{\vec{x} \cdot \vec{v}}{\| \vec{v} \|^2}\right)\vec{v} \end{aligned}$$

If we redefine our line L as all the scalar multiples of out unit vector $\hat{u}$:

$$L = \{c\hat{u} \mid c \in \mathbb{R}\}$$

Simplifies to:

$$\boxed{(\vec{x} \cdot \hat{u})\hat{u}}$$



Projection as a Linear Transformation

Let $\hat{u} \in \mathbb{R}^n$ be a unit vector:

$$\hat{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \quad \text{where } \|\hat{u}\| = 1$$

The projection matrix A is:

$$A = \hat{u}\hat{u}^T = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$$

Expands to:

$$A = \begin{bmatrix} u_1 u_1 & u_1 u_2 & \dots & u_1 u_n \\ u_2 u_1 & u_2 u_2 & \dots & u_2 u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n u_1 & u_n u_2 & \dots & u_n u_n \end{bmatrix}$$

For any $\vec{x} \in \mathbb{R}^n$, the projection of $\vec{x}$ onto the line spanned by $\hat{u}$ is:

$$\text{Proj}_L(\vec{x}) = A\vec{x} = (\hat{u} \cdot \vec{x})\hat{u}$$

Consider a vector $\vec{v}$ in $\mathbb{R}^2$:

$$\vec{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

1. Construct the Unit Vector

$$\|\vec{v} = \sqrt{2^2 + 1^2} = \sqrt{5}$$

$$\hat{u} = \frac{\vec{v}}{\| \vec{v} \|} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$$

This unit vector $\hat{u}$ defines the line L, which consists of all the scalar multiples og $\vec{v}$:

$$L = \{c\vec{v} \mid c \in \mathbb{R}\}$$

2. Derive the Projection Matrix

The projection of any vector $\vec{x}$ onto the line L is given by:

$$\text{Proj}_L(\vec{x}) = A\vec{x}$$

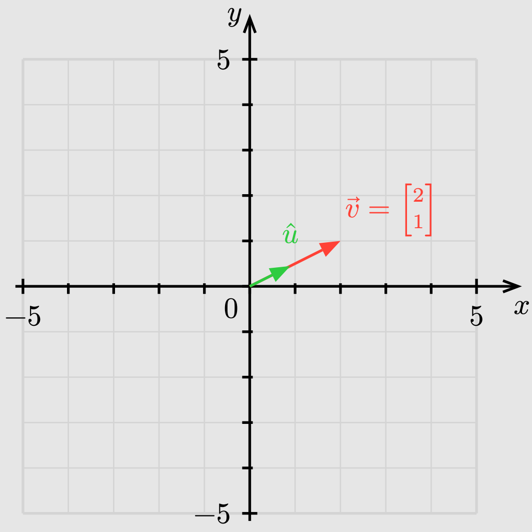
Where A is the projection matrix. To construct A we use the formula:

$$\begin{aligned} A &= \hat{u}\hat{u}^T \\ &= \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\begin{bmatrix} u_1 & u_2 \end{bmatrix} \\ &= \begin{bmatrix} u_1 \cdot u_1 & u_1 \cdot u_2 \\ u_2 \cdot u_1 & u_2 \cdot u_2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}\begin{bmatrix} \dfrac{2}{\sqrt{5}} & \dfrac{1}{\sqrt{5}} \end{bmatrix} \\ &= \begin{bmatrix} \left(\frac{2}{\sqrt{5}}\right)^2 & \frac{1}{\sqrt{5}}\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}}\frac{1}{\sqrt{5}} & \left(\frac{1}{\sqrt{5}}\right)^2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix} \end{aligned}$$

3. Applying the Projection

To project any vector $\vec{x}$ onto L, we multiply $\vec{x}$ by the matrix A:

$$\begin{aligned} \text{Proj}_L(\vec{x}) &= A\vec{x} \\ &= \begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix}\vec{x} \end{aligned}$$



1. Additivity of Projections (Linearity with respect to addition)

$$\begin{aligned}
\mathrm{Proj}_L\left(\vec{a} + \vec{b}\right) &= \left(\left(\vec{a} + \vec{b}\right) \cdot \hat{u}\right) \\
&= \left(\vec{a} \cdot \hat{u} + \vec{b} \cdot \hat{u}\right)\hat{u} \\
&= (\vec{a} \cdot \hat{u})\hat{u} + \left(\vec{b} \cdot \hat{u}\right)\hat{u} \\
&= \mathrm{Proj}_L(\vec{a}) + \mathrm{Proj}_L\left(\vec{b}\right)
\end{aligned}$$

2. Homogeneity of Projections (Linearity with respect to scalar multiplication)

$$\begin{aligned}
\mathrm{Proj}_L(c\vec{a}) &= (c\vec{a} \cdot \hat{u})\hat{u} \\
&= c(\vec{a} \cdot \hat{u})\hat{u} \\
&= c\mathrm{Proj}_L(\vec{a})
\end{aligned}$$

General Properties of A

1. Idempotence

$$A^2 = A$$

2. Symmetry

$$A^T = A$$

3. Rank

$$\mathrm{rank}(A) = 1$$

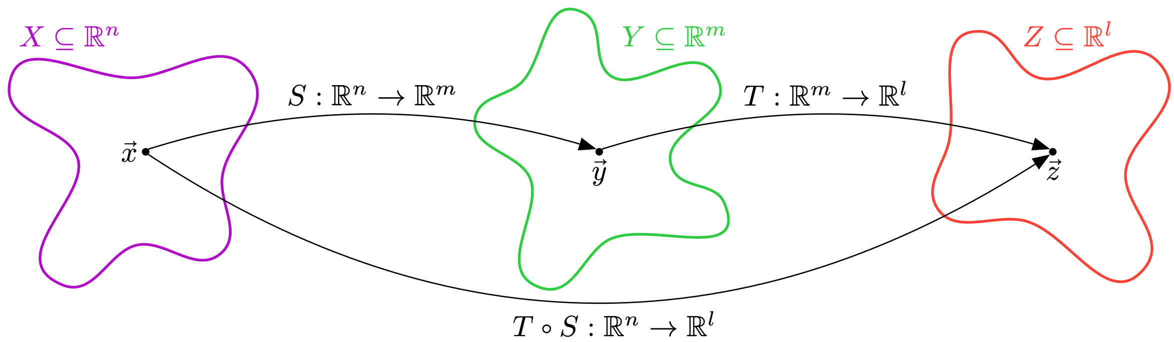
Because $\hat{u}\hat{u}^T$ projects onto a one-dimensional subspace spanned by $\hat{u}$

12.14. Composition of Linear Transformations

$$S : {\color{purple}X} \to {\color{green}Y} \qquad T : {\color{green}Y} \to {\color{red}Z}$$

$$T \circ S : {\color{purple}X} \to {\color{red}Z}$$

$$S(\vec{x}) = \underbrace{A}_{m \times n} \vec{x} \qquad T(\vec{x}) = \underbrace{B}_{l \times m} \vec{x}$$



$$\begin{aligned} T \circ S(\vec{x}) &= T(S(\vec{x})) \\ &= A(B(\vec{x})) \\ &= AB\vec{x} \end{aligned}$$

Consider two linear transformations T and S, where:

- T maps $\mathbb{R}^m \to \mathbb{R}^l$
- S maps $\mathbb{R}^n \to \mathbb{R}^m$

The composition $T \circ S$ is a nre transformation mapping $\mathbb{R}^n \to \mathbb{R}^l$ defined by:

$$T \circ S(\vec{x}) = T(S(\vec{x}))$$

Key Properties of $T \circ S$

1. Additivity

$$\begin{aligned} T \circ S(\vec{x} + \vec{y}) &= T(S(\vec{x} + \vec{y})) \\ &= T(S(\vec{x}) + S(\vec{y})) \\ &= T(S(\vec{x})) + T(S(\vec{y})) \\ &= T \circ S(\vec{x}) + T \circ S(\vec{y}) \end{aligned}$$

2. Homogeneity

$$\begin{aligned} T \circ S(c\vec{x}) &= T(S(c\vec{x})) \\ &= T(cS(\vec{x})) \\ &= cT(S(\vec{x})) \\ &= c(T \circ S)(\vec{x}) \end{aligned}$$

Matrix Representation of $T \circ S$

Let S be represented by the matrix A $(m \times n)$, and let T be respesented by the matrix B $(l \times m)$

For a vector $\vec{x} \in \mathbb{R}^n$

$$\begin{aligned} T \circ S(\vec{x}) &= T(S(\vec{x})) = T(A\vec{x}) \\ &= \underbrace{B}_{l \times m}\left(\underbrace{A}_{m \times n}\vec{x}\right) \\ &= \underbrace{C}_{l \times n}\vec{x} \end{aligned}$$

The composition $T \circ S$ is therefor represented by the matrix $C = A \cdot B$, where C is of size $l \times n$

Column-Wise Interpretation

The matrix A can be decomposed comlumn-wise:

$$A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \end{bmatrix}$$

where $\vec{a}_i$ is the i-th column of A and I_n is the identity matrix in $\mathbb{R}^n$, and its columns are the standard basis vectors $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$.

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \underbrace{0}_{e_1} & \underbrace{0}_{e_2} & \dots & \underbrace{1}_{e_n} \end{bmatrix}$$

To compute C:

1. For each $\vec{e}_i$ in the basis of $\mathbb{R}^n$, $A\vec{e}_i = \vec{a}_i$, the i-th column of A

$$\begin{aligned} C &= \begin{bmatrix} B(Ae_1) & B(Ae_2) & \dots & B(Ae_n) \end{bmatrix} \\ &= \begin{bmatrix} B\left(A\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}\right) & B\left(A\begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}\right) & \dots & B\left(A\begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}\right) \end{bmatrix} \\ &= \begin{bmatrix} B\vec{a}_1 & B\vec{a}_2 & \dots & B\vec{a}_n \end{bmatrix} \end{aligned}$$

The composition $T \circ S$ is the linear map respresented by $\boldsymbol{C} = \boldsymbol{B} \cdot \boldsymbol{A}$

Each column of C reflects how T transforms the action of S on a standard basis vector

12.15. Matrix Product

$$\underbrace{A}_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \qquad \underbrace{B}_{n\times p} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{bmatrix}$$

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

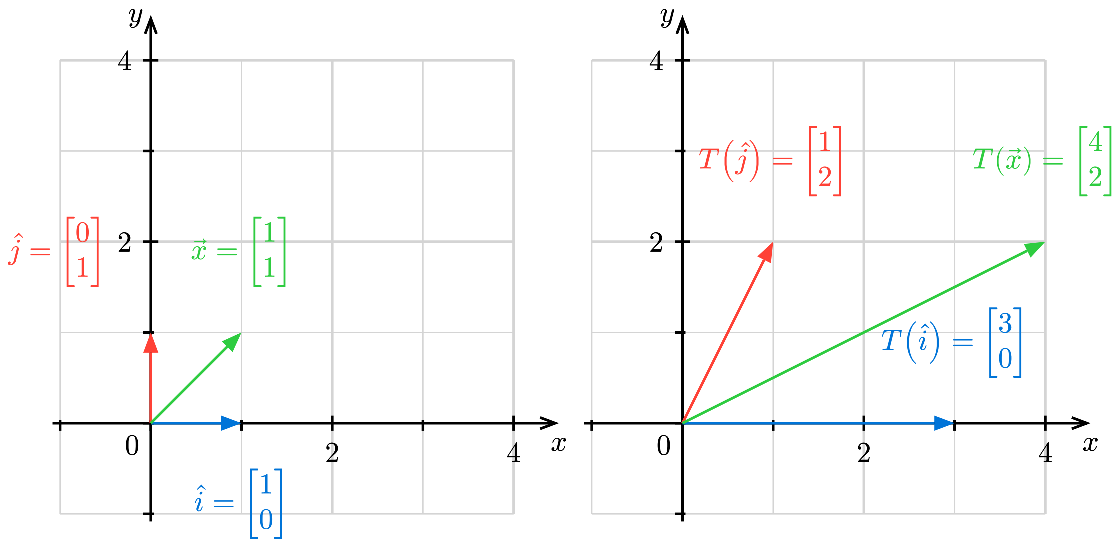
$$\underbrace{A}_{2\times 3} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \qquad \underbrace{B}_{3\times 2} = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}$$

$$
\begin{aligned}
AB &= \begin{bmatrix} A\begin{bmatrix}7\\9\\11\end{bmatrix} & A\begin{bmatrix}8\\10\\12\end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} \begin{bmatrix}1\\2\\3\end{bmatrix}\begin{bmatrix}7\\9\\11\end{bmatrix} & \begin{bmatrix}1\\2\\3\end{bmatrix}\begin{bmatrix}8\\10\\12\end{bmatrix} \\ \\ \begin{bmatrix}4\\5\\6\end{bmatrix}\begin{bmatrix}7\\9\\11\end{bmatrix} & \begin{bmatrix}4\\5\\6\end{bmatrix}\begin{bmatrix}8\\10\\12\end{bmatrix} \end{bmatrix}
\end{aligned}
$$

12.16. Matrix Product Associativity

12.17. Eigenvectors

12.17.1. Transformation



$$
A = \begin{bmatrix}3 & 1\\0 & 2\end{bmatrix}
$$

$$
\begin{aligned}
T\left(\hat{i}\right) &= A\hat{i} \\
&= \begin{bmatrix}3 & 1\\0 & 2\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} \\
&= \begin{bmatrix}3\\0\end{bmatrix}
\end{aligned}
$$

$$T(\hat{j}) = A\hat{j}$$
$$= \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$T(\vec{x}) = A\vec{x}$$
$$= \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$= \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

12.18. Eigenvalues

12.18.1. LU Decomposition

Given a matrix A, LU decomposition aims to express A as:

$$A = LU$$

Where:

- L: Lower triangular matrix (all elements above the diagonal are zero)
- U: Upper triangular matrix (all elements below the diagonal are zero)

1. Solve $Ly = b$ Using Forward Substitution

$$Ly = b$$

Where:

- L: Lower triangular matrix (all elements above the diagonal are zero)
- y: Intermediate vector we are solving for
- b: Right-hand side vector

$$\begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

- First row: $L_{11} y_1 = b_1$, so $y_1 = \frac{b_1}{L_{11}}$
- Second row: $L_{21} y_1 + L_{22} y_2 = b_2$, substitute y_1 into this equation and solve for y_2:

$$y_2 = \frac{b_2 - L_{21} y_1}{L_{22}}$$

- Third row: $L_{31} y_1 + L_{32} y_2 + L_{33} y_2 = b_3$, substitute y_1 and y_2 into this equation, solve for y_3:

$$y_3 = \frac{b_2 - L_{31} y_1 - L_{32} y_2}{L_{33}}$$

2. Solve $Ux = y$ Using Backward Substitution

$$Ux = y$$

Where:

- U: Upper triangular matrix (all elements below the diagonal are zero)
- y: Vector of unknowns (solution)
- b: Vector computed from the forward substitution step

$$\begin{pmatrix} U_{11} & U_{12} & U_{13} \\ 0 & U_{22} & U_{23} \\ 0 & 0 & U_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

- Third row: $U_{33}x_3 = y_3$, so $x_3 = \frac{y_3}{U_{33}}$
- Second row: $U_{22}x_2 + U_{23}x_3 = y_2$, substitute x_3 from the previous step and solve for x_2:

$$x_2 = \frac{y_2 - U_{23}x_3}{U_{22}}$$

- First row: $U_{11}x_1 + U_{12}x_2 + U_{13}x_3 = y_1$, substitute x_2 and x_3 from the previous step and solve for x_1:

$$x_1 = \frac{y_1 - U_{12}x_2 - U_{13}x_3}{U_{11}}$$

$$A = \begin{pmatrix} 2 & 3 & 1 \\ 4 & 7 & 3 \\ 6 & 18 & 5 \end{pmatrix} \qquad b = \begin{pmatrix} 5 \\ 12 \\ 31 \end{pmatrix}$$

1. Factor A into L and U:

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 6 & 1 \end{pmatrix} \qquad U = \begin{pmatrix} 2 & 3 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -2 \end{pmatrix}$$

2. Solve $Ax = b$

   - $y_1 = \frac{b_1}{L_{11}} = \frac{5}{1} = 5$
   - $y_2 = \frac{b_2 - L_{21}y_1}{L_{22}} = \frac{12 - 2\times5}{1} = 2$
   - $y_2 = \frac{b_3 - L_{31}y_1 - L_{32}y_2}{L_{33}} = \frac{31 - 4\times5 - 3\times2}{1} = 5$

So,

$$y = \begin{pmatrix} 5 \\ 2 \\ 5 \end{pmatrix}$$

3. Solve $Ux = y$

   - $x_3 = \frac{y_3}{U_{33}} = \frac{5}{-}2 = -2.5$
   - $x_2 = \frac{y_2 - U_{23}x_3}{U_{22}} = \frac{2 - 1\times-2.5}{1} = 4.5$
   - $x_1 = \frac{y_1 - U_{12}x_2 - U_{13}x_3}{U_{11}} = \frac{5 - 3\times4.5 - 1\times-2.5}{2} = -4$

So,

$$x = \begin{pmatrix} -4 \\ 4.5 \\ -2.5 \end{pmatrix}$$

12.19. Solving Systems of Linear Equations

Linear Equation

$$y = a_1 x_1 + a_1 x_2 + ... + a_n x_n$$

1. Consistency

Whether a system of linear equations has at least one solution

Consistent System

$$x + y = 3$$
$$x - y = 1$$

This system has a unique solution

$$(x, y) = (2, 1)$$

Inconsistent System

$$x + y = 3$$
$$x + y = 5$$

This system is inconsistent (equations contradict each other, no solution can satisfy both)

2. Independence

Whether the equations in the system provide unique and non-redundant information about the variables

Independent Equations

$$x + y = 3$$
$$x - y = 1$$

Neither equation can be derived from the other (they provide unique information and intersect at a single point)

Dependent Equations

$$x + y = 3$$
$$2x + 2y = 6$$

Second equation is just a multiple of the first equation (they describe the same line)

3. Recognizing Systems with No Solution or Infinite Solutions

$$3x + 2y = 6 \qquad \text{(Equation 1)}$$
$$6x + 4y = 12 \qquad \text{(Equation 2)}$$

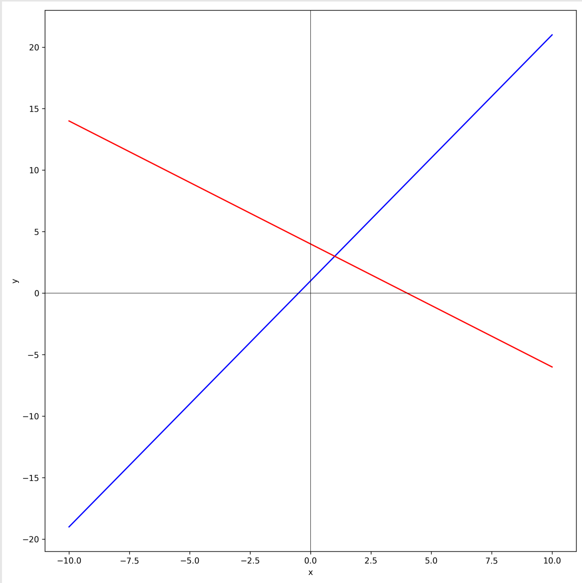


Figure 1: Unique Solution (Consistent and Independent)

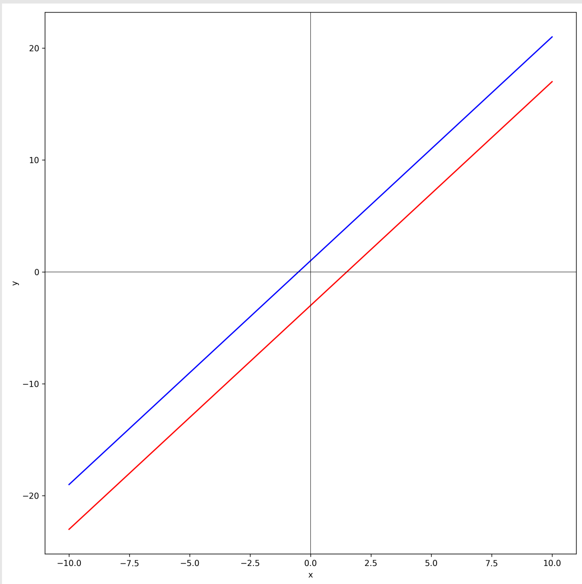


Figure 2: No Solution (Inconsistent)

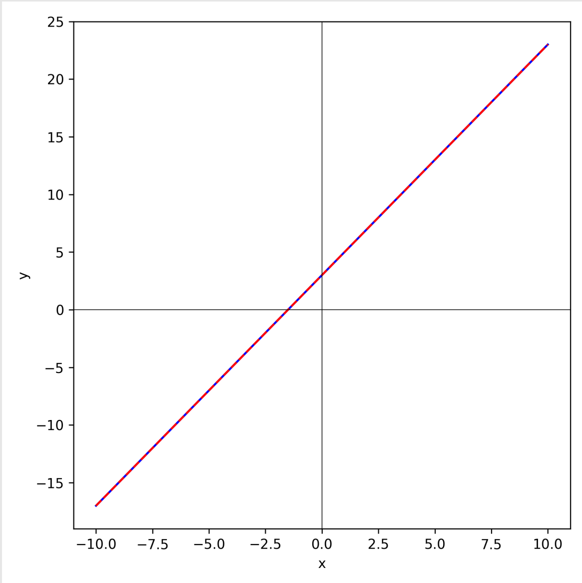


Figure 3: Infinitely Many Solutions (Consistent and Dependent)

2. Matrix Respresentation

System of Equations

$$
\begin{gathered}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\
\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m
\end{gathered}
$$

Matrix Respresentation

Coefficient vector (A)

$$
A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{2m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}
$$

Variable vector (x)

$$
x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}
$$

Constant vector (b)

$$
b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}
$$

Matrix equation

$$
Ax = b
$$

```python
from scipy.linalg import solve

X = np.array([
  [1, 1, 1],
  [2, -1, 3],
  [3, 4, -1]
])

Y = np.array([6, 14, 1])

intersection_point = solve(X, Y)
```

12.19.1. Gaussian Elimination

Convert a matrix into its row echelon form (REF) or reduced row echelon form (RREF)

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

1. Create an augmented matrix

$$A = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & b_3 \\ a_{2m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}\right)$$

2. Forward Elimination

Eliminate the element in the i-th of the k-th column ($k > i$)

$$R_k \leftarrow R_k - \frac{a_{ki}}{a_{ii}} R_i$$

Where

- a_{ii}: Pivot element
- R_k: k-th row
- R_i: i-th row

3. Back Substitution

4. Reduced Row Echelon Form (RREF)

$$\begin{aligned} 2x_1 + 3x_2 &= 5 \\ 4x_1 + 5x_2 &= 5 \end{aligned}$$

1. Create an augmented matrix

$$\left(\begin{array}{cc|c} 2 & 3 & 5 \\ 4 & 5 & 6 \end{array}\right)$$

2. Forward Elimination

$R_k \leftarrow R_k - \frac{a_{ki}}{a_{ii}} R_i$

$R_k \leftarrow R_k - \frac{a_{21}}{a_{11}} R_i$

$R_2 \leftarrow R_2 - \frac{4}{2} R_1$

$R_2 \leftarrow R_2 - 2 \times R_1$

$$\left(\begin{array}{cc|c} 2 & 3 & 5 \\ 4 - 2 \times 2 & 5 - 2 \times 3 & 6 - 2 \times 5 \end{array}\right)$$

Simplifies to:

$$\left(\begin{array}{cc|c} 2 & 3 & 5 \\ 0 & -1 & -4 \end{array}\right)$$

System is now:

$$\begin{aligned} 2x_1 + 3x_2 &= 5 \\ -1x_2 &= -4 \end{aligned}$$

3. Back Substitution

$$-1x_2 = -4$$
$$x_2 = 4$$

Substitute:

$$2x_1 + 3(4) = 5$$
$$2x_1 + 12 = 5$$
$$x_1 = -3.5$$

Solution:

$$x_1 = -3.5$$
$$x_2 = 4$$

12.19.2. Substitution

$$\begin{aligned} x + y &= 10 \qquad \text{(Equation 1)} \\ 2x - y &= 5 \qquad \text{(Equation 2)} \end{aligned}$$

1. Solve Equation 1 for y

$$y = 10 - x$$

2. Substitute into Equation 2

$$2x - (10 - x) = 5$$

3. Solve for x:

$$2x - 10 + x = 5$$
$$3x - 10 = 5$$
$$3x = 15$$
$$x = 5$$

4. Find y using value of x

$$y = 10 - x$$
$$y = 10 - 5$$
$$y = 5$$

12.19.3. Addition or Subtraction Method

$$3x + 2y = 12 \qquad \text{(Equation 1)}$$
$$2x - 2y = 4 \qquad \text{(Equation 2)}$$

1. Add the equations

$$(3x + 2y) + (2x - 2y) = 12 + 4$$
$$5x = 16$$
$$x = \frac{16}{5}$$
$$x = 3.2$$

2. Substitute

$$3(3.2) + 2y = 12$$
$$9.6 + 2y = 12$$
$$2y = 12 - 9.6y = \frac{2.4}{2}$$
$$y = 1.2$$

Calculus I

13. Cheatsheet

13.1. Limits

Epsilon-Delta

For every distance ε around L, there's a δ-range around a that keeps $f(x)$ within ε of L.

$$\lim_{x \to x_0} f(x) = L \iff \forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } 0 < |x - x_0| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

The limit of $f(x)$ as x approaches x_0 equals L if and only if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that, whenever $0 < |x - x_0| < \delta$, implies that $|f(x) - L| < \varepsilon$

Limit Type	**Name**	**Quantifiers**
$x \to x_0, f(x) \to L$	Epsilon-Delta	$\forall \varepsilon, \exists \delta$
$x \to x_0, f(x) \to \infty$	M-Delta	$\forall M, \exists \delta$
$x \to \infty, f(x) \to L$	epsilon-N	$\forall \varepsilon, \exists N$
$x \to \infty, f(x) \to \infty$	M-N	$\forall M, \exists N$

Finite $\to$ Finite (**Epsilon-Delta**)

$$\lim_{x \to x_0} f(x) = L \iff \forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } 0 < |x - x_0| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

Finite $\to$ Infinity (**M-Delta**)

$+\infty$

$$\lim_{x \to x_0} f(x) = +\infty \iff \forall M > 0, \exists \delta > 0 \text{ s.t. } 0 < |x - x_0| < \delta \Rightarrow f(x) > M$$

$-\infty$

$$\lim_{x \to x_0} f(x) = -\infty \iff \forall M > 0, \exists \delta > 0 \text{ s.t. } 0 < |x - x_0| < \delta \Rightarrow f(x) < -M$$

Infinity $\to$ Finite (**Epsilon-N**)

$+\infty$

$$\lim_{x \to +\infty} f(x) = L \iff \forall \varepsilon > 0, \exists N > 0 \text{ s.t. } x > N \Rightarrow |f(x) - L| < \varepsilon$$

$-\infty$

$$\lim_{x \to +\infty} f(x) = L \iff \forall \varepsilon > 0, \exists N > 0 \text{ s.t. } x < -N \Rightarrow |f(x) - L| < \varepsilon$$

Infinity $\to$ Infinity (**M-N**)

$$\lim_{x \to \pm\infty} f(x) = \pm\infty \iff \forall M > 0, \exists N > 0 \text{ s.t. } x < N \Rightarrow f(x) > M$$

13.2. Derivatives

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

Let $f(x) = x^2$

a. Find $f'(x)$

$$f'(x) = \boxed{\text{2x}}$$

b. Prove a

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

c. Prove b

$$\lim_{x \to x_0} f(x) = L \Longleftrightarrow \forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } 0 < |x - x_0| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

p.f.:

Let $\varepsilon > 0$

Choose $\delta = \varepsilon$

Suppose $0 < |h - 0| < \delta$

Check

$$\left| \frac{(x+h)^2 - x^2}{h} - 2x \right|$$

$$
\begin{aligned}
&= \left| \frac{\cancel{x^2} + 2xh + h^2 - \cancel{x^2}}{h} - 2x \right| \\
&= \left| \frac{\cancel{h}(2x + h)}{\cancel{h}} - 2x \right| \\
&= |\cancel{2x} + h - \cancel{2x}| \\
&= \boxed{|h| < \delta = \varepsilon}
\end{aligned}
$$

Constant Rule	$\frac{d}{dx}[c] = 0$	$\int c \, dx = cx + C$
Power Rule	$\frac{d}{dx}[x^n] = nx^{n-1}$	$\int x^n \, dx = \frac{x^{n+1}}{n+1} + C \quad (n \neq -1)$
Constant Multiple Rule	$\frac{d}{dx}[cf] = cf'$	$\int cf dx = c \int f \, dx$
Sum / Difference Rules	$\frac{d}{dx}[f \pm g] = f' \pm g'$	$\int (f \pm g) \, dx = \int f \, dx \pm \int g \, dx$
Product Rule	$\frac{d}{dx}[fg] = f'g + fg'$	
Quotient Rule	$\frac{d}{dx}\left[\frac{f}{g}\right] = \frac{f'g - fg'}{g^2}$	
Chain Rule	$\frac{d}{dx}[f(g(x))] = f'(g(x)) \cdot g'(x)$	
Exponential Function	$\frac{d}{dx}[e^x] = e^x$	$\int e^x dx = e^x + C$
	$\frac{d}{dx}[a^x] = a^x \ln(a)$	$\int a^x dx = \frac{a^x}{\ln a} + C$
Logarithmic Function	$\frac{d}{dx}[\ln(x)] = \frac{1}{x}$	
	$\frac{d}{dx}[\log_a(x)] = \frac{1}{x \ln(a)}$	
Sin	$\frac{d}{dx}[\sin(x)] = \cos(x)$	$\int \cos x \, dx = \sin x + C$

Cos	$\frac{d}{dx}[\cos(x)] = -\sin(x)$	$\int \sin x \, \mathrm{d}x = -\cos x + C$
Tan	$\frac{d}{dx}[\tan(x)] = \sec^2(x)$	$\int \sec^2 x \, \mathrm{d}x = \tan x + C$

14. Limits & Continuity

15. Properties of Limits

15.1. Continuous

15.1.1. Addition, Substraction, Multiplication, Division

$$\lim_{x \to c} (f(x) * g(x)) = \lim_{x \to c} f(x) * \lim_{x \to c} g(x)$$
$$* \in \{+, -, \times, \div\}$$

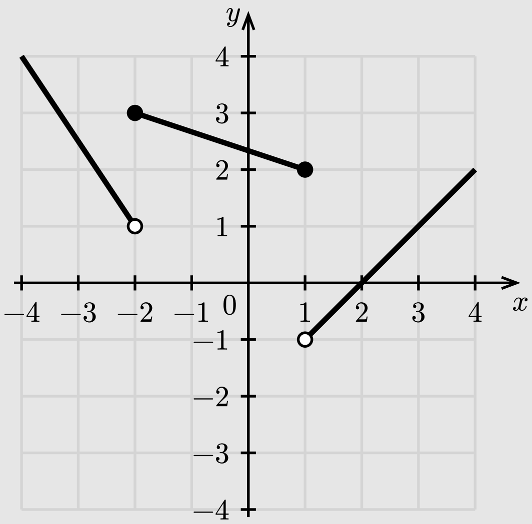
15.1.2. Constant

$$\lim_{x \to c} k f(x) = k \lim_{x \to c} f(x)$$

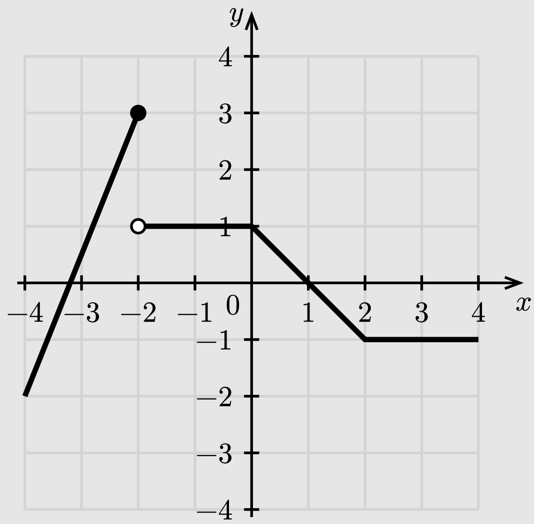
15.2. Non-continuous

Even though the limit for either function may not exist, their $*$ can exist as long as

$$\lim_{x \to c^-} (f(x) * g(x)) \quad = \quad \lim_{x \to c^+} (f(x) * g(x))$$
$$* \in \{+, -, \times, \div\}$$



$f(x)$



$g(x)$

Problem 1: Limit Exists

$$\lim_{x \to -2} (f(x) + g(x))$$

1. Left-hand limit ($x \to -2^-$)

$$\lim_{x\to -2^-} f(x) = 1 \qquad \lim_{x\to -2^-} g(x) = 3$$

Adding these

$$\begin{aligned}\lim_{x\to -2^-}(f(x)+g(x)) &= \lim_{x\to -2^-} f(x) + \lim_{x\to -2^-} g(x)\\ &= 1+3\\ &= 4\end{aligned}$$

2. Right-hand limit ($x \to -2^+$)

$$\lim_{x\to -2^+} f(x) = 3 \qquad \lim_{x\to -2^+} g(x) = 1$$

Adding these

$$\begin{aligned}\lim_{x\to -2^+}(f(x)+g(x)) &= \lim_{x\to -2^+} f(x) + \lim_{x\to -2^+} g(x)\\ &= 3+1\\ &= 4\end{aligned}$$

3. Since both the left-hand and right-hand limits agree

$$\lim_{x\to -2}(f(x)+g(x)) = 4$$

Problem 2: Limit Does Not Exist

$$\lim_{x\to 1}(f(x)+g(x))$$

1. Left-hand limit ($x \to 1^-$)

$$\lim_{x\to 1^-} f(x) = 2 \qquad \lim_{x\to 1^-} g(x) = 0$$

Adding these

$$\begin{aligned}\lim_{x\to 1^-}(f(x)+g(x)) &= \lim_{x\to 1^-} f(x) + \lim_{x\to 1^-} g(x)\\ &= 2+0\\ &= 2\end{aligned}$$

2. Right-hand limit ($x \to 1^+$)

$$\lim_{x\to 1^+} f(x) = -1 \qquad \lim_{x\to 1^+} g(x) = 0$$

Adding these

$$\begin{aligned}\lim_{x\to 1^+}(f(x)+g(x)) &= \lim_{x\to 1^+} f(x) + \lim_{x\to 1^+} g(x)\\ &= -1+0\\ &= -1\end{aligned}$$

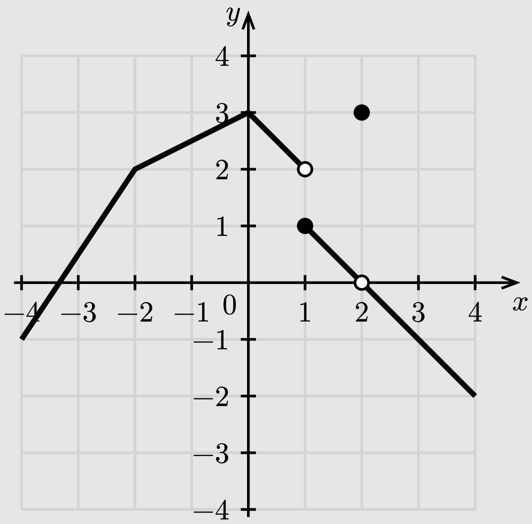
3. Since both the left-hand and right-hand limits do not agree, the limit $\lim_{x\to c}(f(x)+g(x))$ does not exist

15.3. Composite Functions

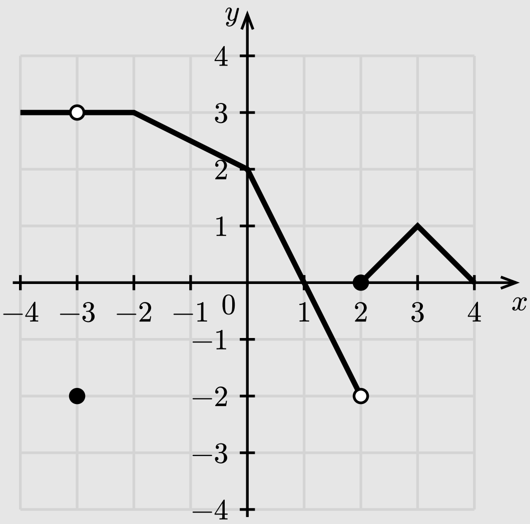
$$\lim_{x \to c} f(g(x)) = f\left(\lim_{x \to c} g(x)\right)$$

For this to hold true, two important conditions must be satisfied:

- **Inner limit exists**: The limit $\lim_{x \to c}$ must exist and equal some value L. That is, as x gets arbitrarily close to c, $g(x)$ approaches a well-defined number L
- **Continuity of the outer function**: The function f must be continuous at the point L. Continuity ensures that f behaves predictably near L, without any jumps, gaps, or undefined points



$f(x)$



$g(x)$

Problem 1: Inner Limit & Continuity Exist

$$\lim_{x \to -3} f(g(x))$$

1. Inner limit $\lim_{x \to -3} g(x)$

Observing $g(x)$, as $x \to -3$, $g(x) \to 3$. The inner limit $L = 3$ exists

$$\lim_{x \to -3} g(x) = 3$$

2. Continuity of $f(x)$ at $x = 3$

Observing $f(x)$, $f(3) = -1$. Since $f(x)$ is continuous at $x = 3$, the composite limit holds

$$\begin{aligned} f\left(\lim_{x \to -3} g(x)\right) &= f(3) \\ &= -1 \end{aligned}$$

Problem 2: Inner Limit Does Not Exist

$$\lim_{x \to 2} f(g(x))$$

1. Inner limit $\lim_{x \to 2} g(x)$

Observing $g(x)$, as $x \to 2$, $g(x) \to 2$. The inner limit does not exist

Problem 3: Continuity Does Not Exist

$$\lim_{x \to 0.5} f(g(x))$$

1. Inner limit $\lim_{x \to 0.5} g(x)$

Observing $g(x)$, as $x \to 0.5$, $g(x) \to 1$. The inner limit $L = 1$ exists

$$\lim_{x \to 0.5} g(x) = 1$$

2. Continuity of $f(x)$ at $x = 1$

Observing $f(x)$, $f(1)$ is not continuous. Since $f(x)$ is not continuous at $x = 1$, the composite limit does not hold

15.4. Limits by Direct Substitution

Limit exists

$$\begin{aligned}\lim_{x \to -1} (6x^2 + 5x - 1) &= 6(-1)^2 + 5(-1) - 1 \\ &= 6 - 5 - 1 \\ &= 0\end{aligned}$$

Limit does not exist (Undefined)

$$\begin{aligned}\lim_{x \to 1} \frac{x}{\ln(x)} &= \frac{1}{\ln(1)} \\ &= \frac{1}{0}\end{aligned}$$

15.4.1. Limits of Piecewise Functions

15.4.2. Absolute Value

15.5. Limits by Factoring

15.6. Limits by Rationalizing

15.7. Continuity & Differentiability at a Point

Piecewise function:

$$f(x) = \begin{cases} x^2 & \text{if } x < 3 \\ 6x - 9 & \text{if } x \geq 3 \end{cases}$$

1. Check for **Continuity**

   - Value of $f(3)$

$$f(3) = 6(3) - 9 = \boxed{9}$$

   - Left-Hand Limit (LHL)

$$\lim_{x \to 3^-} f(x) = 3^2 = \boxed{9}$$

- Right-Hand Limit (RHL)

$$\lim_{x \to 3^+} f(x) = 6(3) - 9 = \boxed{9}$$

Since $f(3) = \lim_{x \to 3^-} f(x) = \lim_{x \to 3^+} f(x)$, $f(x)$ is **continuous** at $x = 3$

2. Check for **Differentiability**

- Left-Hand Derivative (LHD)

$$\begin{aligned}\lim_{x \to 3^-} \frac{f(x) - f(3)}{x - 3} &= \frac{x^2 - 3^2}{x - 3} \\ &= \frac{x^2 - 9}{x - 3} \\ &= \frac{(x+3)(x-3)}{x - 3} \\ &= x + 3 \\ &= \boxed{6}\end{aligned}$$

- Right-Hand Derivative

$$\begin{aligned}\lim_{x \to 3^+} \frac{f(x) - f(3)}{x - 3} &= \frac{(6x - 9) - 3^2}{x - 3} \\ &= \frac{6x - 9 - 9}{x - 3} \\ &= \frac{6x - 18}{x - 3} \\ &= \frac{6(x - 3)}{x - 3} \\ &= \boxed{6}\end{aligned}$$

Since the left-hand and right-hand derivatives are equal, $f(x)$ is **differentiable** at $x = 3$

Conclusion: $f(x)$ is both continuous & differentiable at $x = 3$

15.8. Power Rule

$$f(x) = x^n, \qquad n \neq 0$$
$$f'(x) = nx^{n-1}$$

$$f(x) = x^3$$
$$f'(x) = 3x^2$$

$$\begin{aligned}\frac{d}{dx}\left(\sqrt[3]{x^2}\right) &= \frac{d}{dx}\left(\left(x^2\right)^{\frac{1}{3}}\right)\\ &= \frac{d}{dx}\left(x^{2\times\frac{1}{3}}\right)\\ &= \frac{d}{dx}\left(x^{\frac{2}{3}}\right)\\ &= \frac{d}{dx}\left(\frac{2}{3}x^{-\frac{1}{3}}\right)\end{aligned}$$

15.9. Constant Rule

$$\frac{d}{dx}[k] = 0$$

$$\frac{d}{dx}[-3] = 0$$

15.10. Constant Multiple Rule

$$\begin{aligned}\frac{d}{dx}[kf(x)] &= k\frac{d}{dx}[f(x)]\\ &= kf'(x)\end{aligned}$$

$$\begin{aligned}\frac{d}{dx}[2x^5] &= 2\frac{d}{dx}[x^5]\\ &= 2\cdot 5x^4\\ &= 10x^4\end{aligned}$$

15.11. Sum Rule

$$\begin{aligned}\frac{d}{dx}[f(x)+g(x)] &= \frac{d}{dx}[f(x)] + \frac{d}{dx}[g(x)]\\ &= f'(x) + g'(x)\end{aligned}$$

$$\begin{aligned}\frac{d}{dx}[x^3 + x^{-4}] &= \frac{d}{dx}[x^3] + \frac{d}{dx}[x^{-4}]\\ &= 3x^2 + \left(-4x^{-5}\right)\\ &= 3x^2 - 4x^{-5}\end{aligned}$$

15.12. Difference Rule

$$\begin{aligned}\frac{d}{dx}[f(x)-g(x)] &= \frac{d}{dx}[f(x)] - \frac{d}{dx}[g(x)]\\ &= f'(x) - g'(x)\end{aligned}$$

$$\begin{aligned}\frac{d}{dx}[x^4 - x^3] &= \frac{d}{dx}[x^4] - \frac{d}{dx}[x^3]\\ &= 4x^3 - 3x^2\end{aligned}$$

15.13. Square Root

$$\begin{aligned}
\frac{d}{dx}\sqrt[4]{x} &= \frac{d}{dx}x^{\frac{1}{4}} \\
&= \frac{1}{4}x^{\frac{1}{4}-1} \\
&= \frac{1}{4} \cdot x^{-\frac{3}{4}} \\
&= \frac{1}{4} \cdot \frac{1}{x^{3/4}} \\
&= \frac{1}{4x^{3/4}}
\end{aligned}$$

15.14. Derivative of a Polynomial

$$\begin{aligned}
f(x) &= \color{red}{2x^3} \color{blue}{- 7x^2} + \color{green}{3x} - \color{gold}{100} \\
f'(x) &= \color{red}{2 \cdot 3x^2} \color{blue}{- 7 \cdot 2x} + \color{green}{3} + \color{gold}{0}
\end{aligned}$$

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

Evaluate $\frac{d}{dx}h(x)$ at $x = 9$

$$\begin{aligned}
\frac{d}{dx}(h(x)) &= \frac{d}{dx}(3f(x) + 2g(x)) \\
&= \frac{d}{dx}3f(x) + \frac{d}{dx}2g(x) \\
&= 3\frac{d}{dx}f(x) + 2\frac{d}{dx}g(x)
\end{aligned}$$

Evaluate $h'(9)$

$$h'(9) = 3f'(9) + 2g'(9)$$

$$\begin{aligned}
g(x) &= \frac{2}{x^3} - \frac{1}{x^2} \\
\frac{d}{dx}(g(x)) &= \frac{d}{dx}(2x^{-3} - 1x^{-2}) \\
g'(x) &= 2 \cdot (-3)x^{-4} - (-2)x^{-3} \\
&= -6x^{-4} + 2x^{-3}
\end{aligned}$$

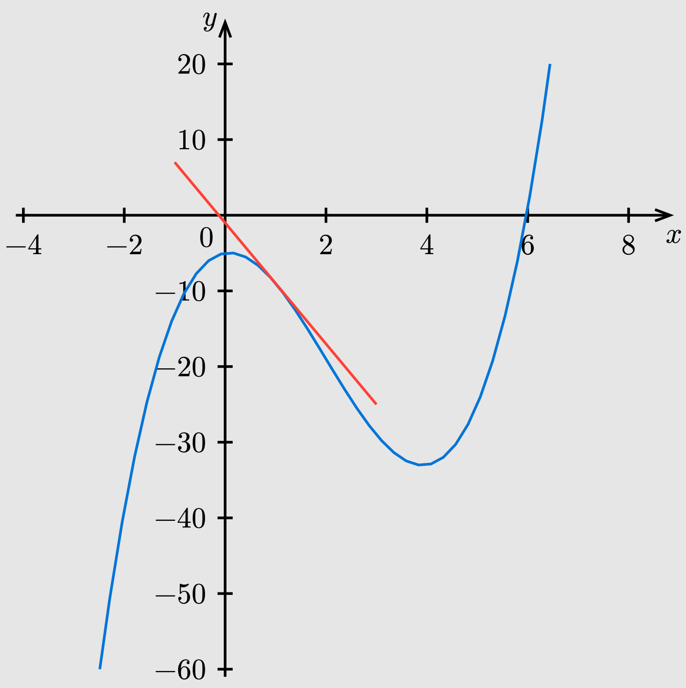
$$
\begin{aligned}
g'(2) &= -6 \cdot 2^{-4} + 2 \cdot 2^{-3} \\
&= -\frac{6}{2^4} + \frac{2}{2^3} \\
&= -\frac{3}{8} + \frac{2}{8} \\
&= -\frac{1}{8}
\end{aligned}
$$

−9 −8

$$
\begin{gathered}
f(x) = x^3 - 6x^2 + x - 5 \\
y = mx + b \\
f'(1) = -8 \; y = -8 \; x + b \\
-9 = -8 \cdot 1 + b \\
-9 = -8 + b \\
-9 + 8 = -8 + 8 + b \\
b = -1
\end{gathered}
$$

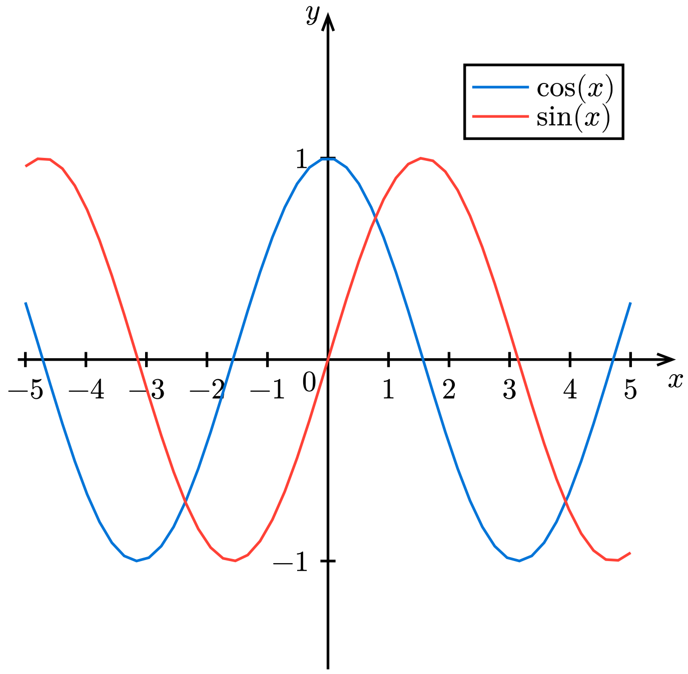
Tangent line to $f(x)$ at $x = 1$

$$y = -8x - 1$$



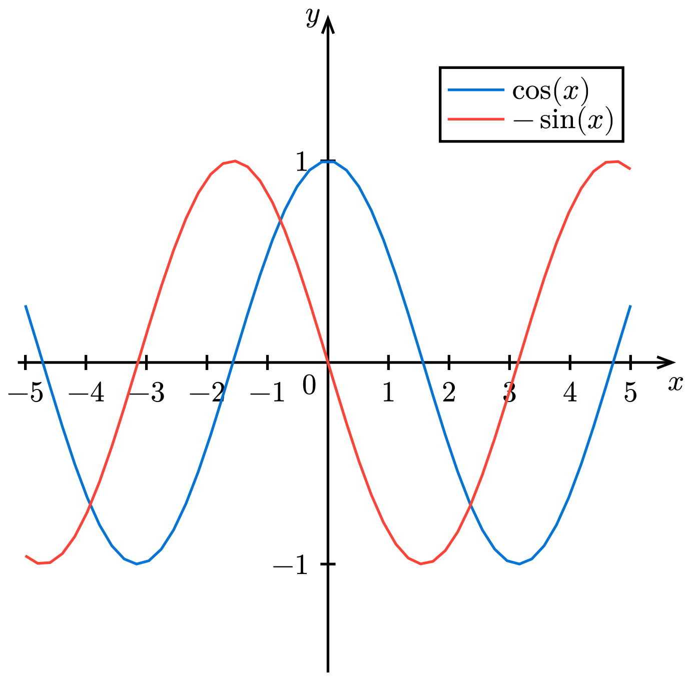
15.15. Sin

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



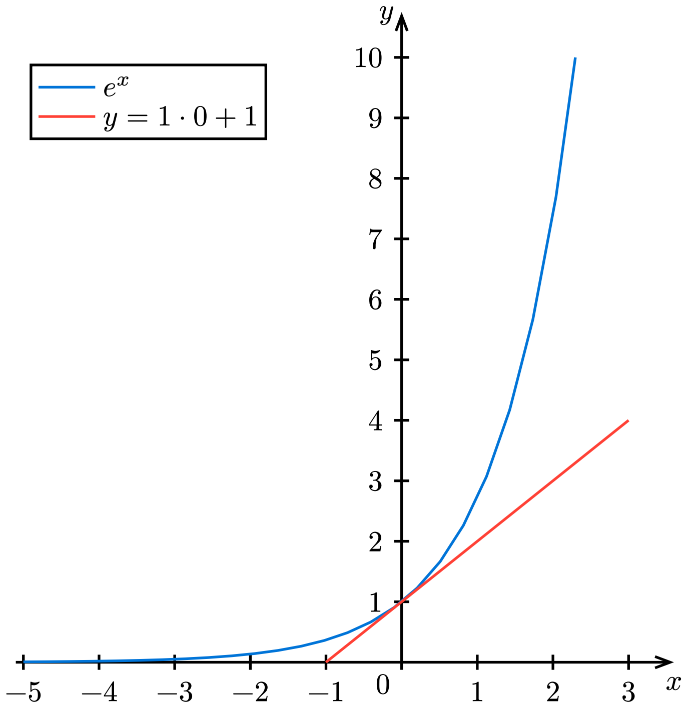
15.16. Cos

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



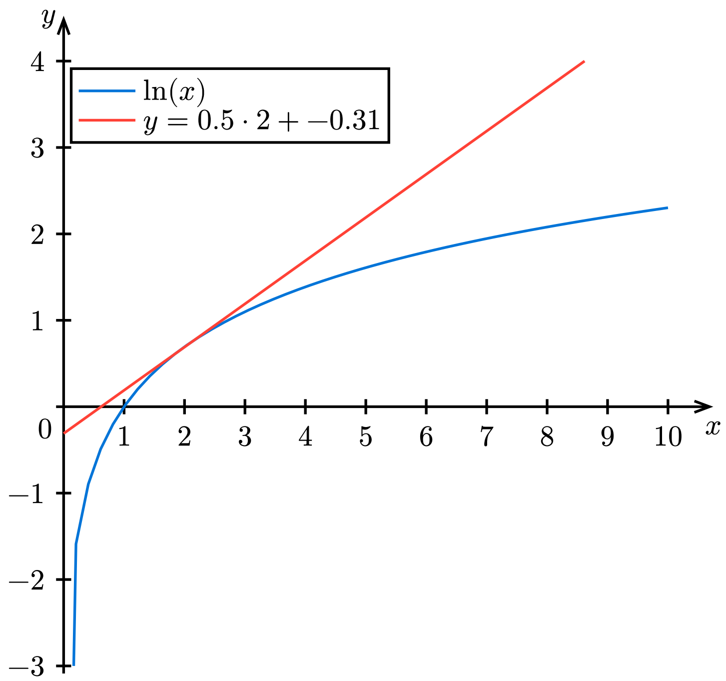
15.17. e^x

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



15.18. $\ln(x)$

$$\ln(x) = \frac{1}{x}$$



15.19. Product Rule

$$\frac{d}{dx}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)$$

$$\frac{d}{dx}[x^2 \sin(x)]$$

$$f(x) = x^2 \qquad g(x) = \sin(x)$$
$$f'(x) = 2x \qquad g'(x) = \cos(x)$$

$$\frac{d}{dx}[x^2 \sin(x)] = 2x \sin(x) + x^2 \cos(x)$$

15.20. Quotient Rule

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$$

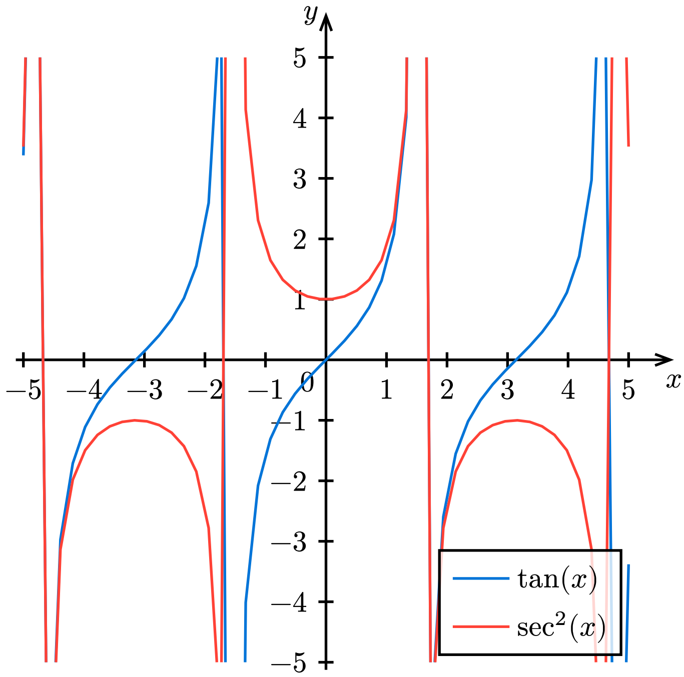
$$\frac{d}{dx}\left[\frac{x^2}{\cos(x)}\right]$$

$$f(x) = x^2 \qquad g(x) = \cos(x)$$
$$f'(x) = 2x \qquad g'(x) = -\sin(x)$$

$$\frac{d}{dx}\left[\frac{x^2}{\cos(x)}\right] = \frac{2x\cos(x) - x^2(-\sin(x))}{[\cos(x)]^2}$$

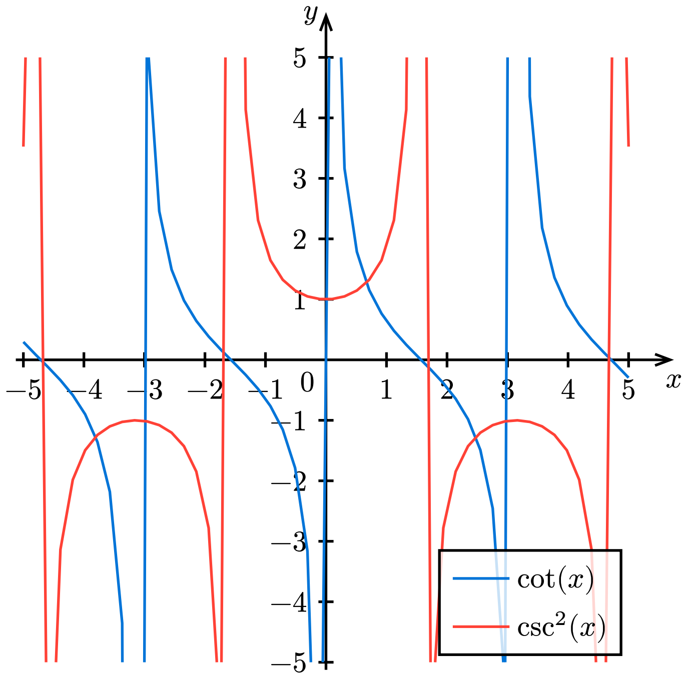
15.20.1. $\tan(x)$

$$\begin{aligned}
\frac{d}{dx}[\tan(x)] &= \frac{d}{dx}\left[\frac{\sin(x)}{\cos(x)}\right] \\
&= \frac{\cos(x)\cdot\cos(x) - \sin(x)\cdot -\sin(x)}{\cos^2(x)} \\
&= \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)} \\
&= \boxed{\tfrac{1}{\cos^2(x)}} \\
&= \boxed{\sec^2(x)}
\end{aligned}$$



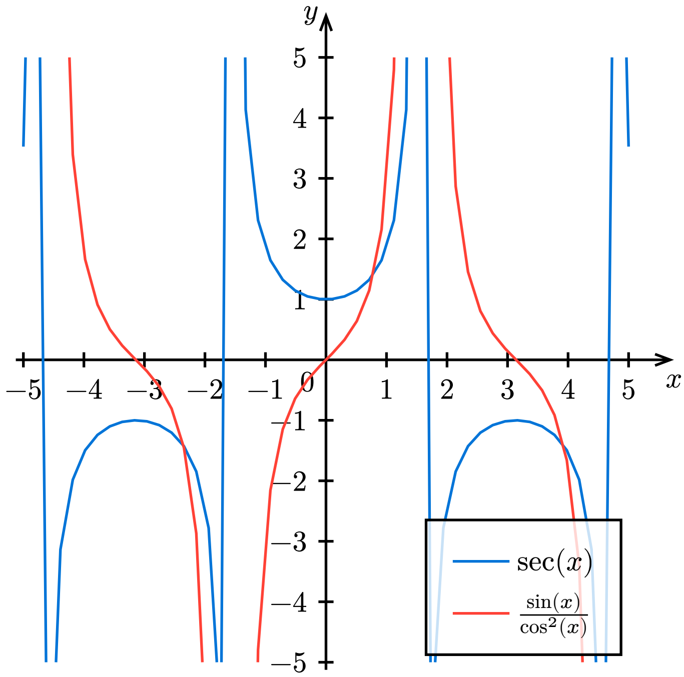
15.20.2. $\cot(x)$

$$\begin{aligned}
\frac{d}{dx}[\cot(x)] &= \frac{d}{dx}\left[\frac{\cos(x)}{\sin(x)}\right] \\
&= \frac{\cos(x)\cdot\cos(x) - \sin(x)\cdot -\sin(x)}{\cos^2(x)} \\
&= \frac{-\sin^2(x) - \cos^2(x)}{} \\
&= \boxed{-\frac{1}{\sin^2}(x)} \\
&= \boxed{-\csc^2(x)}
\end{aligned}$$



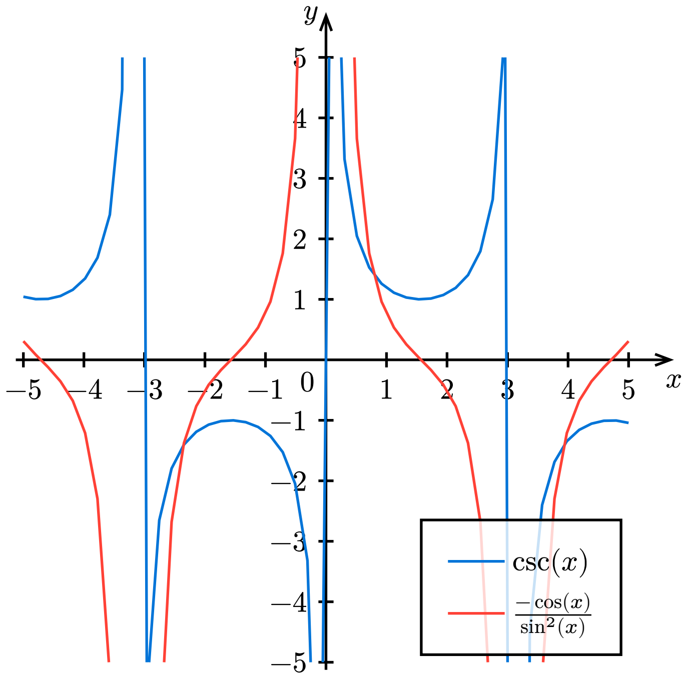
15.20.3. $\sec(x)$

$$\begin{aligned}
\frac{d}{dx}[\sec(x)] &= \frac{d}{dx}\left[\frac{1}{\cos(x)}\right] \\
&= \frac{0\cdot\cos(x) - 1\cdot -\sin(x)}{\cos^2(x)} \\
&= \frac{0 + 1\cdot\sin(x)}{\cos^2(x)} \\
&= \boxed{\frac{\sin(x)}{\cos^2(x)}} \\
&= \frac{\sin(x)}{\cos(x)}\cdot\frac{1}{\cos(x)} \\
&= \boxed{\tan(x)\cdot\sec(x)}
\end{aligned}$$



15.20.4. $\csc(x)$

$$
\begin{aligned}
\frac{d}{dx}[\csc(x)] &= \frac{d}{dx}\left[\frac{1}{\sin(x)}\right] \\
&= \frac{0 \cdot \sin(x) - 1 \cdot \cos(x)}{\sin^2(x)} \\
&= \frac{0 - 1 \cdot \cos(x)}{\sin^2(x)} \\
&= \boxed{\frac{-\cos(x)}{\sin^2(x)}} \\
&= -\frac{\cos(x)}{\sin(x)} \cdot \frac{1}{\sin(x)} \\
&= \boxed{\cot(x) \cdot \csc(x)}
\end{aligned}
$$



15.21. Chain Rule

Compute the derivative of a composite function

$$\begin{aligned} h'(x) &= \frac{d}{dx}[f(g(x))] \\ &= f'(g(x)) \cdot g'(x) \end{aligned}$$

More generally:

$$y = f_1(f_2(f_3(...f_n(x)...)))$$

$$\frac{dy}{dx} = f_1'(f_2(f_3(...f_n(x)...))) \cdot f_2'(f_3(...f_n(x)...)) \cdot f_3'(...f_n(x)...) \cdot ... \cdot f_n'(x)$$

$$\frac{d}{dx}[f(g(x))] = f'(g(x)) \cdot g'(x)$$

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

$$\frac{d}{dx}[e^x] = \frac{1}{x} \qquad \frac{d}{dx}[\sin(x)] = \cos(x)$$

$$f'(g(x)) = \frac{1}{\sin(x)} \qquad g'(x) = \cos(x)$$

$$\frac{d}{dx}\left[\underbrace{\ln\left(\overbrace{\sin(x)}^{g(x)}\right)}_{f(g(x))}\right] = \frac{1}{\sin(x)} \cdot \cos(x)$$

$$f(x) = \cos^3(x) = (\cos(x))^3$$

$$f(x) = v(u(x))$$

$$f'(x) = v'(u(x)) \cdot u'(x)$$

$$\begin{aligned} f'(x) &= \frac{dv}{du} \cdot \frac{du}{dx} \\ &= \frac{d(\cos(x))^3}{d\cos(x)} \cdot \frac{d\cos(x)}{x} \\ &= 3(\cos(x))^2 \cdot -\sin(x) \\ &= -3(\cos(x))^2 \sin(x) \end{aligned}$$

15.22. Implicit Differentiation

When a function is not explicitly solved for one variable in terms of another. E.g.:

$$x^2 + y^2 = 1$$

Instead of solving for y explicitly in terms of x, implicit differentiation allows you to differentiate both sides of an equation directly, treating y as an implicit function of x.

Steps for Implicit Differentiation

1. Differentiate both sides of the equation with respect to x, treating y as a function of x
2. Apply the chain rule whenever differentiating y, since $y = y(x)$
2. Solve for $\frac{dx}{dy}$

$$x^2 + y^2 = 1$$

1. Differentiate both sides

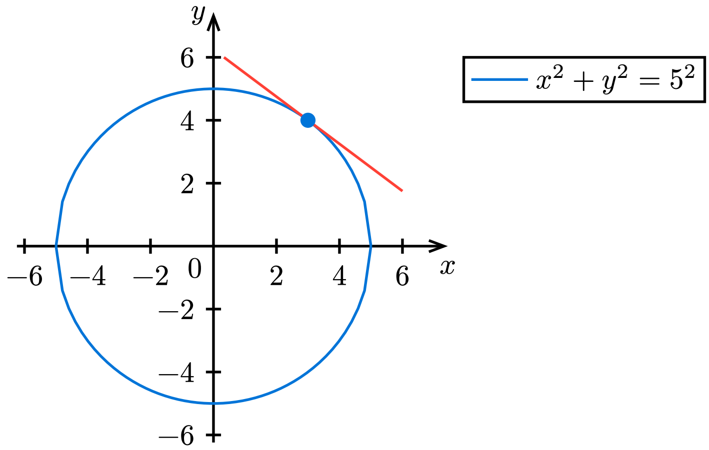
$$\begin{aligned} \frac{d}{dx}[x^2 + y^2] &= \frac{d}{dx}[1] \\ \frac{d}{dx}[x^2] + \frac{d}{dx}[y^2] &= \frac{d}{dx}[1] \\ \frac{d}{dx}[x^2] + \frac{d}{dx}[y^2] &= 0 \end{aligned}$$

2. Apply the chain rule to y^2

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

3. Slove for $\frac{dy}{dx}$

$$\begin{aligned} 2y \cdot \frac{dy}{dx} &= -2x \\ \frac{dy}{dx} &= \frac{-2x}{2y} \\ \frac{dy}{dx} &= -\frac{x}{y} \end{aligned}$$



$$x^2 + y^2 = 5^2$$
$$2xdx + 2ydy = 0$$
$$\frac{dy}{dx} = -\frac{x}{y}$$

15.23. Derivatives of Inverse Functions

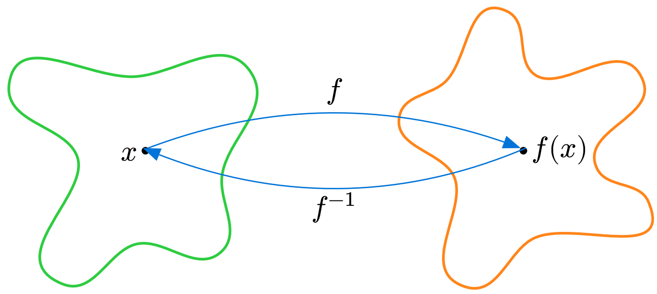
Finding the derivative of the inverse function $f^{-1}(x)$ at a given point directly from the function $f(x)$. Instead of explicitly computing the inverse function $f^{-1}(x)$, we use the inverse function derivative formula:

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

This approach allows us to determine the derivative of the inverse function without needing to express $f^{-1}(x)$ explicitly. Instead, we find the value of x that satisfies $f(x) = a$ (where a a is the given point), evaluate $f'(x)$, and apply the formula.

1. Definition of Inverse Function

$$f(f^{-1}(x)) = x$$
$$f^{-1}(f(x)) = x$$



2. Differentiate Both Sides

Differenatiate both sides with respect to x

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

The right-hand side simplifies to:

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

Using chain-rule:

$$\frac{d}{dx}[f(g(x))] = f'(g(x)) \cdot g'(x)$$

The left side expands as:

$$\frac{d}{dx}[f(f^{-1}(x))] = f'(f^{-1}(x)) \cdot \frac{d}{dx}[f^{-1}(x)]$$

This we get:

$$f'(f^{-1}(x)) \cdot \frac{d}{dx}[f^{-1}(x)] = 1$$

3. Solve for $\frac{d}{dx}[f^{-1}(x)]$

Rearrange to isolate $\frac{d}{dx}[f^{-1}(x)]$:

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

Given the function:

$$f(x) = x^3$$

We want to find $\frac{d}{dx} f^{-1}(x)$ at $x = 0.5$ using the inverse function derivative formula:

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

---

1. Compute $f'(x)$

Differenatiate $f(x)$:

$$f'(x) = 3x^2$$

---

2. Solve for x such that $f(x) = 0.5$

We need to find x such that:

$$x^3 = 0.5$$

Solving for x:

$$x = \sqrt[3]{0.5}$$

---

3. Compute $f'\left(\sqrt[3]{0.5}\right)$

Evaluate the derivative at $x = \sqrt[3]{0.5}$:

$$f'\left(\sqrt[3]{0.5}\right) = 3\left(\sqrt[3]{0.5}\right)^2$$

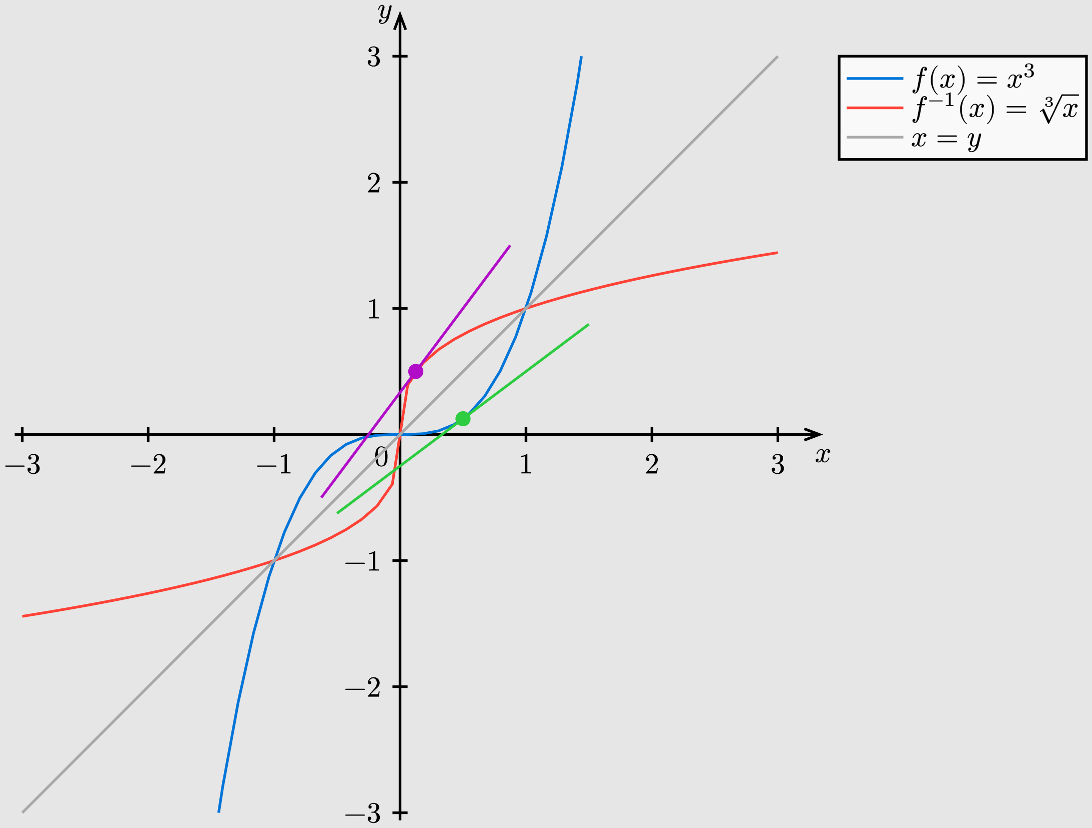
---

4. Use the formula

$$\frac{d}{dx}f^{-1}(0.5) = \frac{1}{3\left(\sqrt[3]{0.5}\right)^2}$$

---

5. Interpretation



The expression:

$$\frac{d}{dx}f^{-1}(0.5) = \frac{1}{3\left(\sqrt[3]{0.5}\right)^2}$$

represents the derivative of the inverse function $f^{-1}(x)$ evaluated at $x = 0.5$. This means it gives the slope of the tangent line to the inverse funcion at $x = 0.5$.

```
from sympy import symbols, solve

# y = x**3 + x

x, y = symbols('x y')

f = x**3 + x - y

inverse = solve(f, x)

print(inverse)
```

15.23.1. Derivative Inverse Sin

$$\frac{d}{dx}[\sin^{-1}(x)] = \frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx}[\arcsin(x)] = \frac{1}{\sqrt{1-x^2}}$$

15.23.2. Derivative Inverse Cos

$$\frac{d}{dx}[\cos^{-1}(x)] = -\frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx}[\arccos(x)] = -\frac{1}{\sqrt{1-x^2}}$$

15.23.3. Derivative Inverse Tan

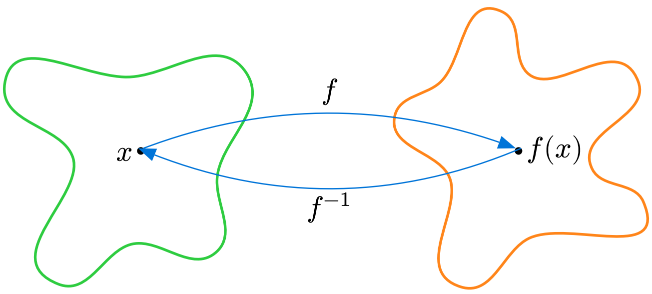
$$\frac{d}{dx}[\tan^{-1}(x)] = -\frac{1}{1+x^2}$$

$$\frac{d}{dx}[\arctan(x)] = -\frac{1}{1+x^2}$$

15.24. Inverse Functions

1. Definition

$$f(f^{-1}(x)) = x$$
$$f^{-1}(f(x)) = x$$



A function $f : A \to B$ has an inverse function f^{-1} if and only if f is **bijective** (i.e., both one-to-one and onto):

- **Injective (One-to-One)**:

  $f(x_1) = f(x_2)$ implies $x_1 = x_2$

  No two inputs map to the same output
- **Surjective (Onto)**:

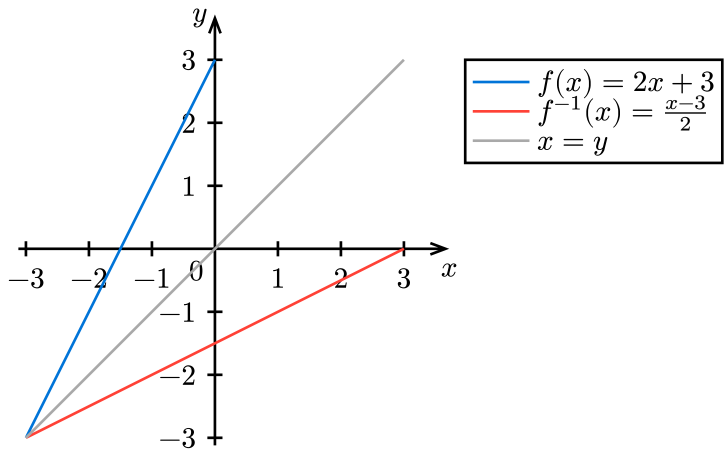
  Every element in B is mapped from some element in A

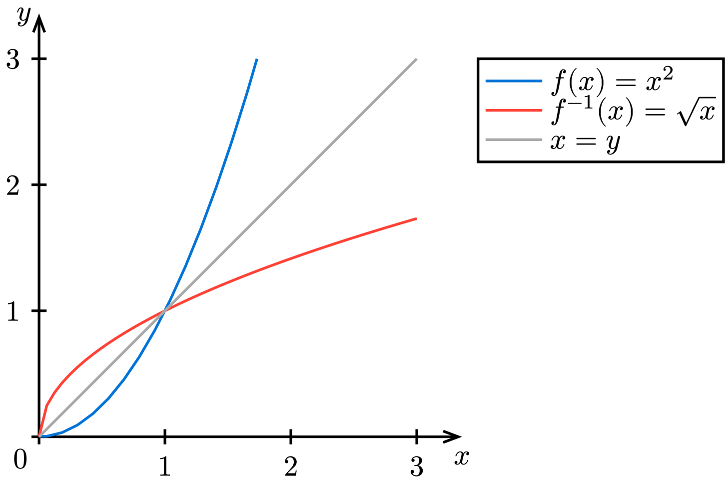
2. Finding Inverse Function

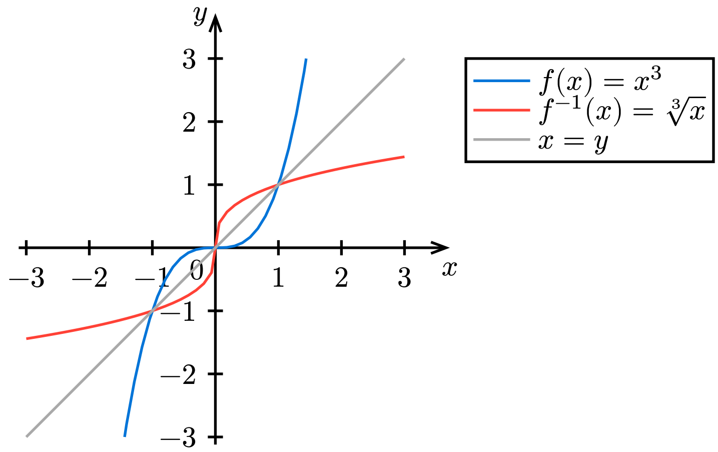
To determine f^{-1}:
- Express y in terms of x: $y = f(x)$
- Solve for x in terms of y
- Swap x and y, remaining y as $f^{-1}(x)$

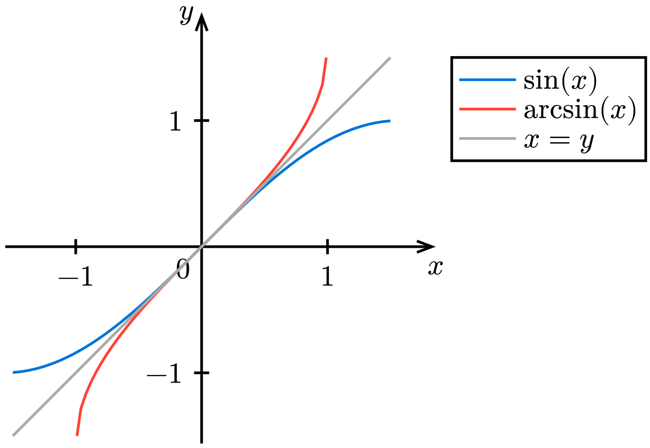
3. Graphical Representation

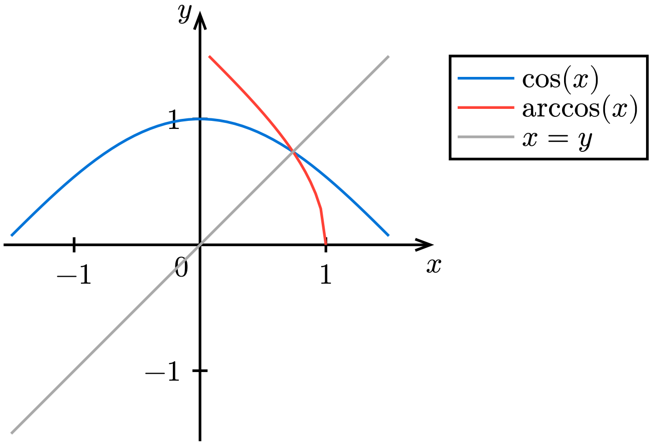
The graph of f^{-1} is a reflection of the graph of f accress the line $x = y$

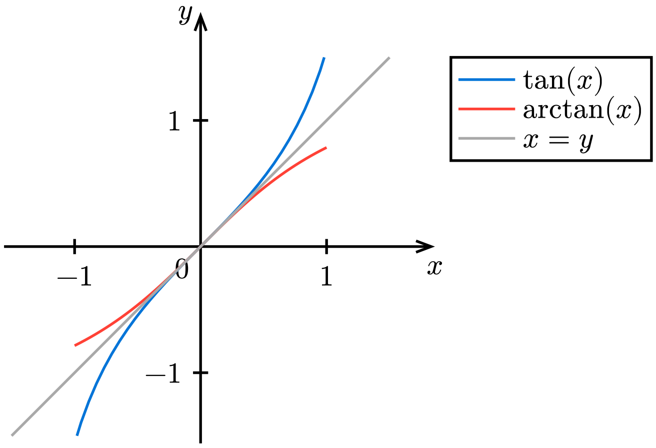












15.25. L'Hôpital's Rule

Evaluating limits that result in a indeterminite form like $\frac{0}{0}$ or $\frac{\infty}{\infty}$

If $\lim_{x \to a} f(x) = 0$ and $\lim_{x \to a} g(x) = 0$ (or both go to $\pm\infty$), and $f(x)$ and $g(x)$ are differentiable near a, then:

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

Consider:

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

Step 1: Direct Substitution

Substituting $x = 0$:

$$\frac{1 - \cos(0)}{0^2} = \frac{1 - 1}{0} = \frac{0}{0}$$

Since this is an indeterminate form, we apply L'Hôpital's Rule.

Step 2: First Application of L'Hôpital's Rule

Differentiate the numerator and denominator:

- Numerator: $f(x) = 1 - \cos(x) \Rightarrow f'(x) = \sin(x)$
- Denominator: $g(x) = x^2 \Rightarrow g'(x) = 2x$

Thus, applying L'Hôpital's Rule:

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

Step 3: Second Check for Indeterminate Form

Substituting $x = 0$:

$$\frac{\sin(0)}{2(0)} = \frac{0}{0}$$

Since this is still an indeterminate form, we apply L'Hôpital's Rule again.

Step 4: Second Application of L'Hôpital's Rule

Differentiate again:

- Numerator: $f'(x) = \sin(x) \Rightarrow f''(x) = \cos(x)$
- Denominator: $g'(x) = 2x \Rightarrow g''(x) = 2$

Applying L'Hôpital's Rule again:

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

Step 5: Evaluate the Limit

Now, substituting $x = 0$:

$$\frac{\cos(0)}{2} = \frac{1}{2}$$

Final Answer:

$$\boxed{\lim_{x \to 0} \frac{1 - \cos(x)}{x^2} = \frac{1}{2}}$$

15.26. Mean Value Theorem

Let $f : [a, b] \to \mathbb{R}$ be a function tha satifies the following conditions:

1. f is **continuous** on the closed interval $[a, b]$
2. f is **differentiable** on the open interval (a, b)

Then there exists at least one point $c \in (a, b)$ such that

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

This means that the instantaneous rate of change (derivative) at some point c is equal to the average rate of change over the entire interval

Consider $f(x) = x^2$ on $[1, 3]$

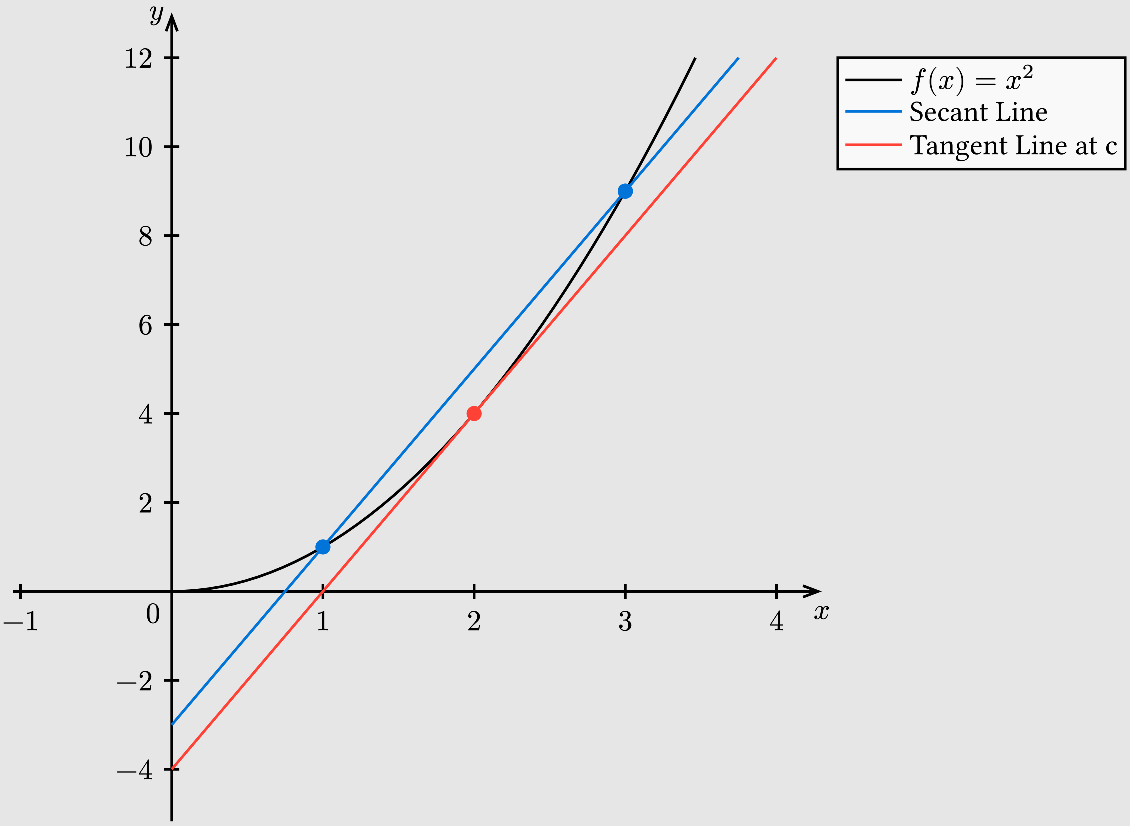
- The average rate of change is:

$$\frac{f(3) - f(1)}{3 - 1} = \frac{9 - 1}{2} = 4$$

- The derivative for $f(x)$ is $f'(x) = 2x$
- Setting $f'(c) = 4$, we solve:

$$2c = 4 \Rightarrow c = 2$$

Thus, at $c = 2$, the instantaneous rate of change matches the average rate of change

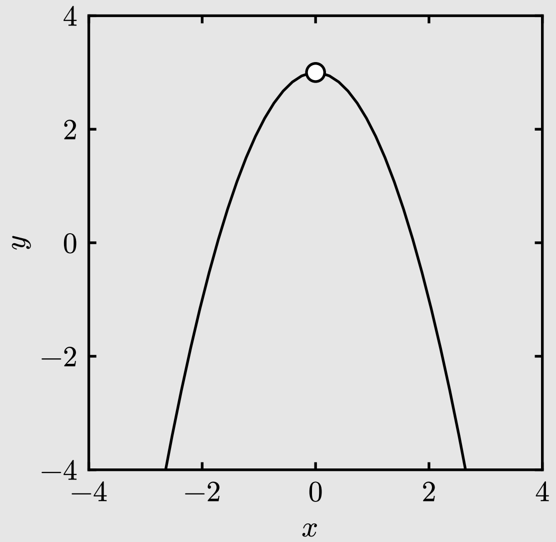


15.27. Extreme Value Theorem

The Extreme Value Theorem states that if a function $f(x)$ is continuous on a closed interval $[a, b]$, then $f(x)$ must attain both a maximum and a minimum value within that interval. This means there exist points $c, d \in [a, b]$ such that:

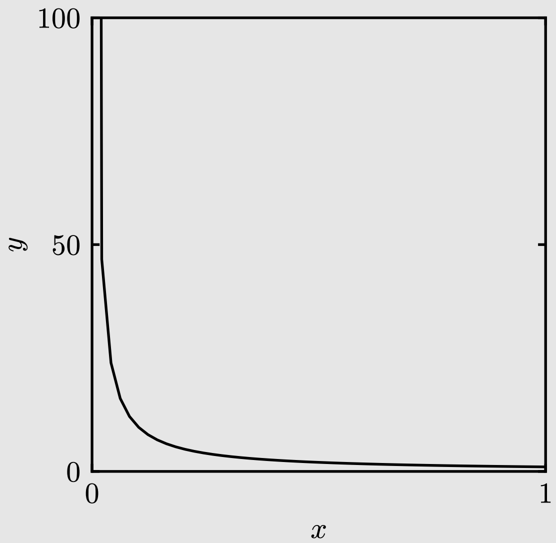
$$f(c) \geq f(x) \quad \text{and} \quad f(d) \leq f(x) \quad \text{for all} \quad x \in [a, b]$$

1. **Continuity**: The function must be continuous on $[a, b]$. Discontinuities (jumps, asymptotes, holes) can prevent the function from attaining an extreme value



2. **Closed Interval**: If the function is defined on an open interval (a, b), an extremum may not exist

$f(x) = \frac{1}{x}$on $(0, 1]$ has no maximum because it keeps increasing as $x \to 0$.

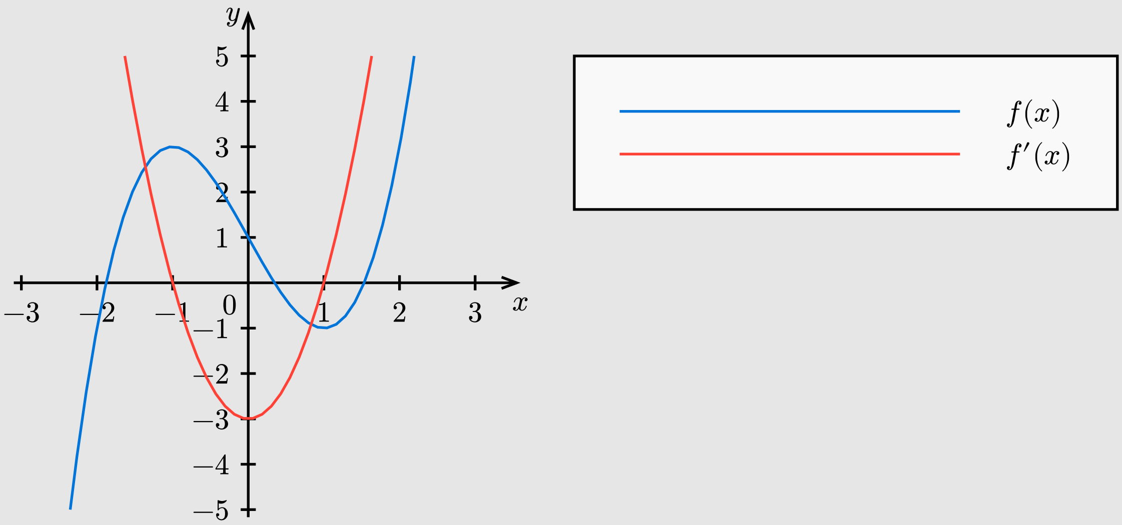


15.27.1. Critical points

A critical point of a function $f(x)$ is a point in the domain where either:

- $f'(x) = 0$
- $f'(x)$ is undefined

$f(x) = \frac{1}{x}$ on $(0, 1]$ has no maximum because it keeps increasing as $x \to 0$.



15.27.2. Global vs. Local Extrema

- $f(c)$ is a **relative maximum** if $f(c) \geq f(x)$ for all $x \in (c - h, c + h)$ for $h > 0$
- $f(d)$ is a **relative minimum** if $f(d) \leq f(x)$ for all $x \in (d - h, d + h)$ for $h > 0$

15.27.3. First and Second Derivative Tests

First Derivative Test and the Second Derivative Test are used to classify critical points, and both aim to determine whether the point is a local maximum, local minimum, or neither

15.27.3.1. First Derivative Test

If $f'(x)$ changes sign around a critical point c, we can determine if $f(c)$ is a local maximum or minimum:

- If $f'(x)$ changes from **positive to negative** at c, then $f(c)$ is a **local maximum**
- If $f'(x)$ changes from **negative to positive** at c, then $f(c)$ is a **local minimum**
- If $f'(x)$does **not** change sign, $f(c)$ is **not** a local extremum

Function:

$$f(x) = x^3 - 3x$$

Derivative:

$$\begin{aligned} f'(x) &= 3x^2 - 3 \\ &= 3(x - 1)(x + 1) \end{aligned}$$

Critical points ($f'(x) = 0$):

- $x = -1$
- $x = 1$

Analysis of $f'(x)$

- Pick point left of -1, say $x = -2$:

$$f'(-2) = 3(4-1) = 9 > 0 \rightarrow \textbf{increasing}$$

- Pick a point between -1 and 1, say $x = 0$:

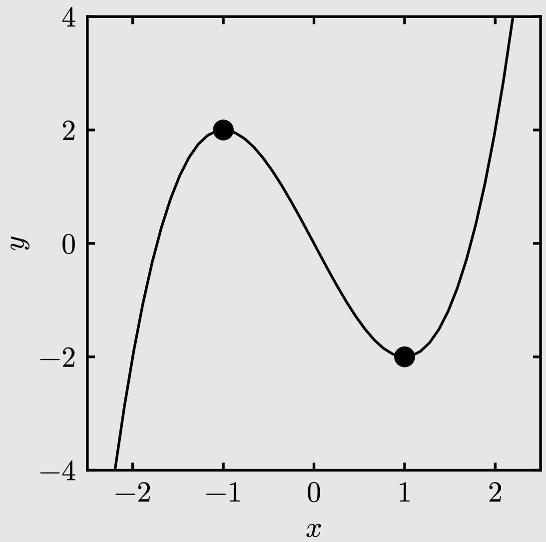
$$f'(0) = -3 < 0 \rightarrow \textbf{decreasing}$$

- Pick a point right of 1, say $x = 2$:

$$f'(2) = 3(4-1) = 9 > 0 \rightarrow \textbf{increasing}$$

Conclusion

- At $x = -1$, $f'(x)$ changes from **positive to negative** $\rightarrow$ **local maximum**
- At $x = 1$, $f'(x)$ changes from **negative to positive** $\rightarrow$ **local minimum**



15.27.3.2. Second Derivative Test

If $f''(x)$ is continuous near a critical point c, and $f'(c) = 0$, then:

- If $f''(c) > 0$, then $f(c)$ is a **local minimum** (concave up)
- If $f''(c) < 0$, then $f(c)$ is a **local maximum** (concave down)
- If $f''(c) = 0$, the test is **inconclusive** — use the first derivative test or other methods

Function:

$$f(x) = x^4 - 4x^2$$

First Derivative:

$$f'(x) = 4x^3 - 8x$$

Second Derivative:

$$f''(x) = 12x^2 - 8$$

Critical points ($f'(x) = 0$):

- $x = 0$
- $x = \sqrt{2}$

- $x = -\sqrt{2}$

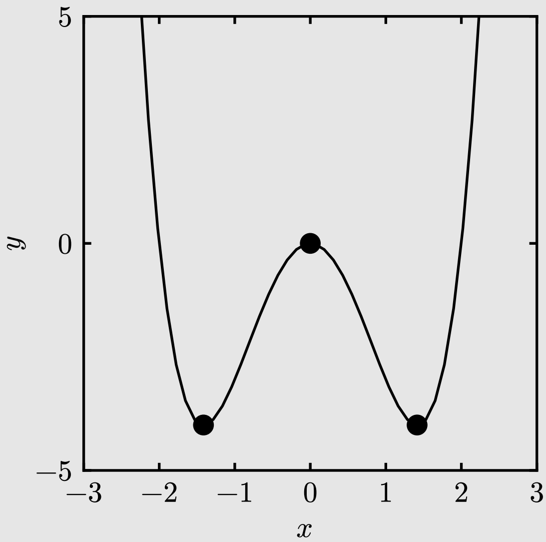
Evaluate $f''(x)$ at each point

- At $x = 0$

  $$f''(0) = 12(0)^2 - 8 = -8 \rightarrow \textbf{concave down} \rightarrow \textbf{local maximum}$$

- At $x = \sqrt{2}$ or $x = -\sqrt{2}$

  $$f''\left(\sqrt{2}\right) = f''\left(-\sqrt{2}\right) = 12(2)^2 - 8 = 16 \rightarrow \textbf{concave up} \rightarrow \textbf{local minimum}$$



15.28. Differentiation Rules

Power	$\frac{d}{dx}[x^n] = n \cdot x^{n-1}$
Sum	$\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$
Product	$\frac{d}{dx}[f(x) \cdot g(x)] = f'(x)g(x) + f(x)g'(x)$
Quotient	$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$
Chain	$\frac{d}{dx}[f(g(x))] = f'(g(x)) \cdot g'(x)$

Calculus II

Calculus III

Statistics

16. Probability Theory

16.1. Probability Axioms

16.1.1. Non-Negativity

Probability of any event cannot be negative

$$P(A) \geq 0$$

16.1.2. Normalization

The probability of the sample space is always 1

$$P(S) = 1$$

Where:

- S is the sample space (the set of all possible outcomes)

16.1.3. Additivity

If two events are mutually exclusive (cannot happen at the same time), the probability of either occurring is the sum of their individual probabilities

$$P(A \cup B) = P(A) + P(B)$$

16.2. Rules

16.2.1. Complement Rule

The probability of the complement of an event A is $P(A^c) = 1 - P(A)$

How likely it is that the event does not occur.

> Consider a fair die:
>
> - Let A be the event "rolling a 4". Then $P(A) = \frac{1}{6}$
> - The complement of A, denoted $P(A^c)$, is "not rolling 4"
> - $P(A^c) = 1 - P(A) = 1 - \frac{1}{6} = \frac{5}{6}$

16.2.2. Multiplication Rule

- Independent Events

If A and B are independent, then $P(A \cap B) = P(A) \times P(B)$. The probability of both events occurring is the product of their individual probabilities.

> Consider flipping two fair coins. Let A be the event "the first coin is heads" and B be the event "the second coin is heads"
>
> - Since the flips are independent, $P(A \cap B) = P(A) \times P(B)$
> - $P(A) = \frac{1}{2}$ and $P(B) = \frac{1}{2}$
> - Thus, $P(A \cap B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

- Dependent Events

If A and B are dependent, $P(A \cap B) = P(A) \times P(B \mid A)$, where $P(B \mid A)$ is the conditional probability of B given A.

Draw two cards from a standard deck without replacement. Let A be the event "drawing an Ace on the first draw" and B be the event "drawing an Ace on the second draw"

- $P(A) = \frac{4}{52} = \frac{1}{13}$
- If A occurs (i.e., an Ace is drawn first), there a re 3 Aces left out of 51 cards. So, $P(B \mid A) = \frac{3}{51} = \frac{1}{17}$
- Thus, $P(A \cap B) = P(A) \times P(B \mid A) = \frac{1}{13} \times \frac{1}{17} = \frac{1}{221}$

16.2.3. Addition Rule

For Any Two Events: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. This accounts for the overlap between the two events to avoid double counting.

Suppose you roll a die, and you want to find the probability of rolling a 2 or a 4.

- Let A be the event "rolling a 2," and B be the event "rolling a 4"
- $P(A) = \frac{1}{6}$ and $P(B) = \frac{1}{6}$
- Since A and B are mutually exclusive, $P(A \cap B) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{1}{6} - 0 = \frac{2}{6} = \frac{1}{3}$

16.2.4. Conditional Probability

The probability of an event A given that B has occurred is $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$, provided $P(B) > 0$.

In a deck of 52 cards, if you know a card is a spade, what is the probability that it is an Ace?

- Let A be the event "drawing an Ace," and B be the event "drawing a spade"
- $P(B) = \frac{12}{52} = \frac{1}{4}$
- There is 1 Ace of Spades out of 13 spades, so $P(A \cap B) = \frac{1}{32}$
- Thus, $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{52}}{\frac{13}{52}} = \frac{1}{13}$

16.2.5. Law of Total Probability

If $\{B_i\}$ is a partition of the sample space, then for any event A:

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A \mid B_i) \times P(B_i)$$

Suppose you want to calculate the probability of raining on a given day. You know that it's either sunny or cloudy, and the probability of rain is different in each condition.

- Let B_1 be the event "sunny" and B_2 be the event "cloudy"
- $P(B_1) = \frac{3}{4}$ and $P(B_2) = \frac{1}{4}$

- The probability of rain given sunny is $P(A \mid B_1) = \frac{1}{10}$, and given cloudy is $P(A \mid B_2) = \frac{2}{5}$
- The total probability of rain is:

$$\begin{aligned} P(A) &= P(A \mid B_1) \times P(B_1) + P(A \mid B_2) \times P(B_2) \\ &= \frac{1}{10} \times \frac{3}{4} + \frac{2}{5} \times \frac{1}{4} \\ &= \frac{3}{40} + \frac{2}{20} \\ &= \frac{3}{40} + \frac{4}{40} \\ &= \frac{7}{40} \end{aligned}$$

16.2.6. Law of Large Numbers

As the number of trials of a random experiment increases, the sample mean will converge to the expected value (mean) of the random variable

You flip a fair coin 1000 times. As you increase the number of flips, the proportion of heads should get closer to the probability of heads (which is 0.5).

16.2.7. Central Limit Theorem

For a sufficiently large number of trials, the distribution of the sample mean approaches a normal distribution, regardless of the distribution of the population

Suppose you are rolling a fair die repeatedly and calculating the average of the results. If you take 30 samples, each being the average of 10 die rolls, the distribution of those sample means will be approximately normal with a mean of 3.5 (the mean of a fair die) and a standard deviation that can be computed based on the original die's distribution.

16.3. Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

- $P(A)$: prior probability (initial belief about event A)
- $P(B \mid A)$: likelihood (probability of observing B given that A is true
- $P(B)$: marginal likelihood (overall probability of observing B under all possible conditions)
- $P(A \mid B)$: posterior probability (updated belief after considering the evidence B

17. Descriptive Statistics

17.1. Central Tendency

17.1.1. Mean

Sum of all the values divided by the number of values

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$[1, 2, 3]$$

$$\overline{x} = \frac{1 + 2 + 3}{3} = 150$$

17.1.2. Median

Middle value in a set of values when they are arranged in ascending or descending order

1. **Odd** Number of Values

$$[1, 2, 3]$$

- **Step 1**: Arrange the Data in Ascending Order

$$[1, 2, 3]$$

- **Step 2**: Identify the Median

$$\text{Median} = 2$$

2. **Even** Number of Values

$$[1, 2, 3, 4]$$

- **Step 1**: Arrange the Data in Ascending Order

$$[1, 2, 3, 4]$$

- **Step 2**: Identify the Median

$$\text{Median} = \frac{2 + 3}{2} = \frac{5}{2} = 2.5$$

17.1.3. Mode

Value that appears most frequently

$$[1, 1, 2, 3]$$

- **Step 1**: Identify the Most Frequent Number

$$\begin{gathered} 1 : 2 \\ 2 : 1 \\ 3 : 1 \end{gathered}$$

- **Step 2**: Determine the Mode

$$\text{Mode} = 1$$

17.2. Dispersion

17.2.1. Range

$$\text{Range} = \max - \min$$

$$[1, 2, 3, 4, 5]$$

- **Step 1**: Identify the Maximum and Minimum Values

$$\max = 5$$
$$\min = 1$$

- **Step 2**: Calculate the Range

$$\text{Range} = 5 - 1 = 4$$

17.2.2. Variance

Quantifies the spread or dispersion of a set of data points in a dataset

- Population

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{n} (x_i - \mu)^2$$

- Sample

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

$$[70, 75, 80, 85, 90]$$

Step 1: Find mean

$$\overline{x} = \frac{70 + 75 + 80 + 85 + 90}{5} = \frac{400}{5} = 80$$

Step 2: Subtract the Mean and Square the result

$$(70 - 80)^2 = (-10)^2 = 100$$
$$(75 - 80)^2 = (-5)^2 = 25$$
$$(80 - 80)^2 = 0^2 = 0$$
$$(85 - 80)^2 = 5^2 = 25$$
$$(90 - 80)^2 = 10^2 = 100$$

Step 3: Calculate variance

$$s^2 = \frac{100 + 25 + 0 + 25 + 100}{5 - 1} = \frac{250}{4} = 62.5$$

17.2.3. Standard deviation

Quantifies the amount of variation or dispersion in a set of data points

- Population

$$\sigma = \sqrt{\sigma^2}$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$$

- Sample

$$s = \sqrt{s^2}$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (x_i - \overline{x})^2}$$

17.2.4. Interquartile Range (IQR)

$$\text{IQR} = \text{Q3} - \text{Q1}$$

$$[1, 2, 3, 4, 5, 6, 7]$$

Step 1: Arrange the Data in Ascending Order

$$[1, 2, 3, 4, 5, 6, 7]$$

Step 2: Find the Quartiles

1. Calculate the Median (Q2)

$$\text{Median } (\text{Q2}) = 4$$

2. Find First Quartile (Q1)

Q1 is the median of the first half of the dataset

$$\text{Q1} = 2$$

3. Find Third Quartile (Q3)

Q3 is the median of the second half of the dataset

$$\text{Q3} = 5$$

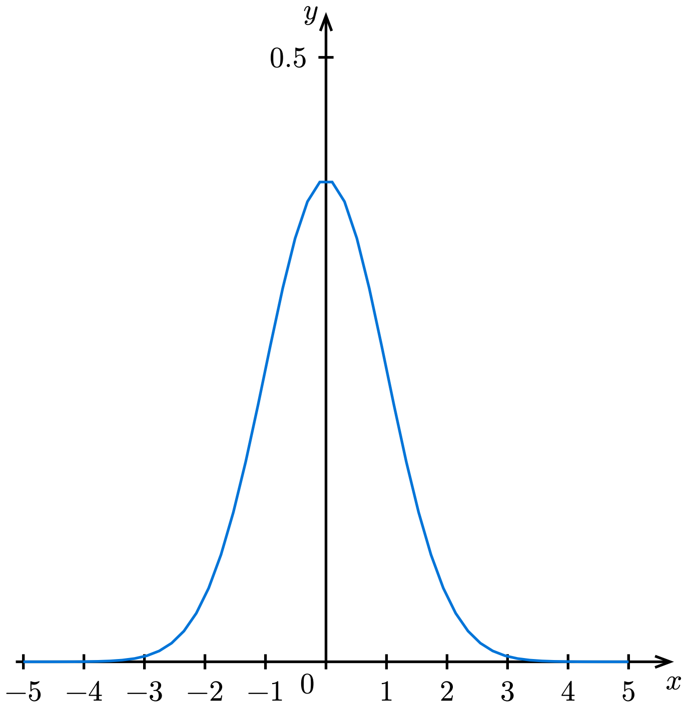
Step 3: Calculate the Interquartile Range (IQR)

$$\text{IQR} = 5 - 2 = 3$$

18. Probability Distributions

18.1. Gaussian (Normal) distribution

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

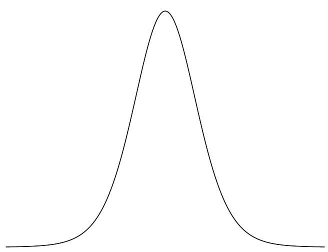


18.2. t-Distribution

$$f(t \mid \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Where:

- t: t-statistic
- ν (or df): degrees of freedom
- Γ: Gamma function (generalizes the factorial function)



Continuous probability distribution for estimating the mean of a normally distributed population in situations where:

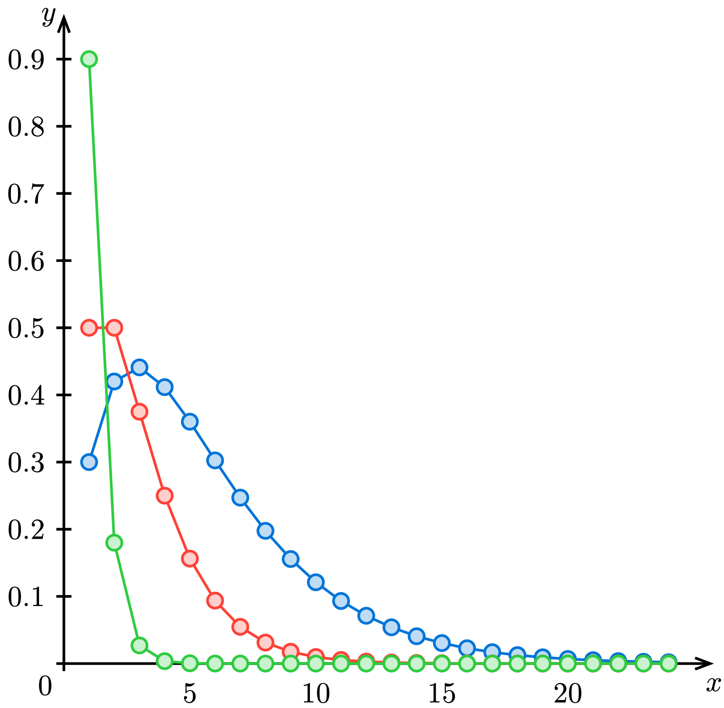
- Sample size is small
- Population standard deviation is unknown

Similar in shape to the normal distribution but has heavier tails, which means it gives more probability to values further from the mean

18.3. Binomial distribution

Discrete probability distribution that describes the number of successes in a fixed number of independent Bernoulli trials, each with the same probability of success

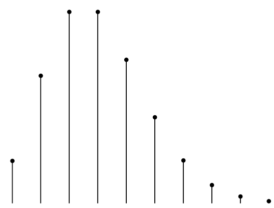
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$



18.4. Poisson distribution

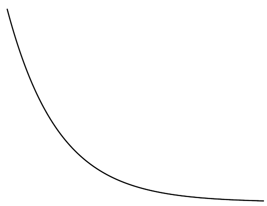
Used to model the number of events that occur within a fixed interval of time or space, given a constant mean rate and assuming that these events occur independently of each other

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$



18.5. Exponential distribution

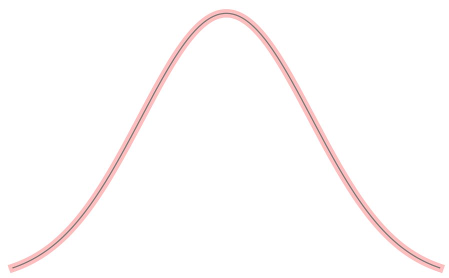
$$f(x \mid \lambda) = \lambda e^{-\lambda x}$$



19. Functions

19.1. PDF (Probability Density Function)

Function that describes the likelihood of a continuous random variable taking on a particular value



Properties:

- The area under the curve of a PDF over the entire range of possible values equals 1
- The PDF itself is non-negative everywhere
- The probability that the variable falls within a certain range is given by the integral of the PDF over that range

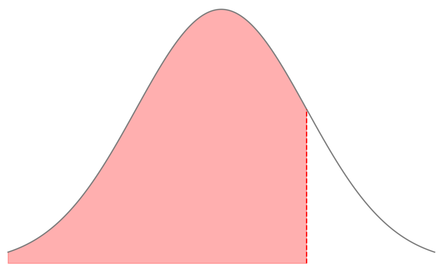
19.2. PMF (Probability Mass Function)

Frequency distribution that provides the probability that a categorical variable takes on each of its possible values

19.3. CDF (Cumulative Distribution Function)

Gives the probability that X will take a value less than or equal to x

$$F(x) = P(X \leq x)$$



1. Categorical

$$F(x) = \int_{-\inf}^{x} f(t) dt$$

2. Continuous

$$F(x) = \sum_{t \leq x} P(X = t)$$

```
from scipy.stats import norm
```

```
x = 1
mu = 0
sigma = 1

norm.cdf(x, loc=mu, scale=sigma)
```

19.4. PPF (Percent-Point Function)

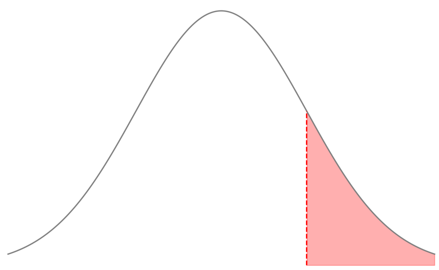
Gives the value x such that the probability of a random variable being less than or equal to x is equal to a given probability p

```
from scipy.stats import norm

p = 0.95
mu = 0
sigma = 1

x_value = norm.ppf(p, loc=mu, scale=sigma)
```

19.5. SF (Survival Function)

Probability that a certain event has not occurred by a certain time

$$S(t) = P(T > t)$$



Relationship to PDF:

$$S(t) = 1 - F(t)$$

```
from scipy.stats import norm

z = 3.4
mu = 0
sigma = 1

norm.sf(z, loc=mu, scale=sigma)
```

20. Error Metrics

20.1. MAE (Mean Absolute Error)

Average of squared differences between predicted $(\hat{y}_i)$ and actual values (y_i)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

20.2. MSE (Mean Squared Error)

Average of squared differences between predicted ($\hat{y}_i$) and actual (y_i) values

$$\mathrm{MSE} = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

20.3. RMSE (Root Mean Squared Error)

square root of the average squared differences between predicted ($\hat{y}_i$) and actual (y_i) values

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

20.4. MAPE (Mean Absolute Percentage Error)

Average percentage difference between predicted ($\hat{y}_i$) and actual (y_i) values

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

20.5. R-squared

Proportion of variance explained by the model

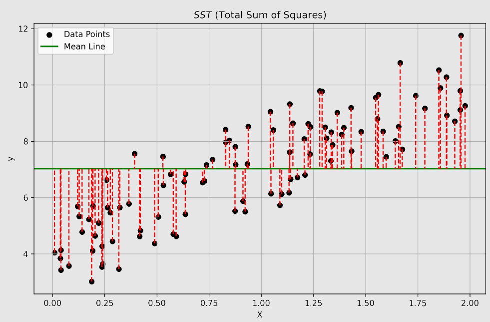
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2}$$

- 1: model explains all the variance in the dependent variable
- 0: model explains none of the variance in the dependent variable

Step 1: Fit the Regression Model

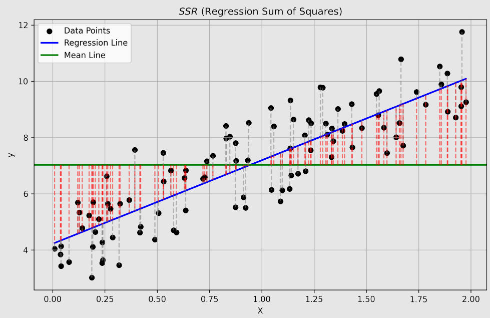
Step 2: Compute **Total Sum of Squares** (SST)

$$\mathrm{SST} = \sum_{i=1}^{n}(y_i - \overline{y})^2$$



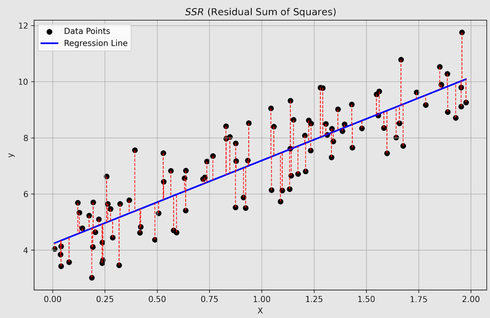
Step 3: Compute **Regression Sum of Squares** (SSR)

$$\text{SSR} = \sum_{i=1}^{n} = (\hat{y}_i - \overline{y})^2$$



Step 4: Compute **Residual Sum of Squares** (SSE)

$$\text{SSE} = \sum_{i=1}^{n} = (y_i - \hat{y}_i)^2$$



Step 5: Calculate $\boldsymbol{R^2}$

$$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$$

20.6. Adj R-squared

Adjusts the R^2 value based on the number of predictors (penalty for adding non-informative variables)

$$\text{Adj } R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

20.7. MSLE (Mean Squared Logarithmic Error)

When predictions and actual values span several orders of magnitude

$$\text{MSLE} = \frac{1}{n}\sum_{i=1}^{n}(\log(1 + y_i) - \log(1 + \hat{y}_i))^2$$

20.8. Cross-Entropy Loss (Log Loss)

Binary and multi-class classification

$$\text{Log Loss} = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

21. Hypothesis Testing

21.1. Hypotheses

21.1.1. Null (H_0)

No effect or difference (observed effect is due to sampling variability)

$$H_0 : \mu = \mu_0$$

21.1.2. Alternative (H_1 or H_a)

Presence of an effect or a difference

- Two-Tailed

$$H_1 : \mu \neq \mu_0$$

- Right-Tailed

$$H_1 : \mu > \mu_0$$

- Left-Tailed

$$H_1 : \mu < \mu_0$$

21.2. Error Types

21.2.1. Type I (α)

$$\alpha = P(\text{Reject } H_0 \mid H_0 \text{ is true})$$

21.2.2. Type II (β)

$$\beta = P(\text{Fail to Reject } H_0 \mid H_1 \text{ is true})$$

21.3. t-Tests

21.3.1. One-sample

Tests if the mean of a single sample differs from a known or hypothesized population mean.

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

- $\overline{x}$: sample mean
- μ_0: hypothesized population mean
- s: sample standard deviation
- n: sample size

$$[78, 82, 89]$$

Step 1: State Hypotheses

$$H_0 : \mu = 85 \text{ cm}$$

$$H_1 : \mu \neq 85 \text{ cm}$$

Step 2: Summarize Data

- Sample values:

$$x_1 = 78, x_2 = 82, x_3 = 89$$

- Sample size:

$$n = 3$$

Step 3: Calculate Sample Mean ($\overline{x}$)

$$\overline{x} = \frac{x_1 + x_2 + x_3}{n} = \frac{78 + 82 + 89}{3} = 83$$

Step 4: Calculate Sample Standard Deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n - 1}}$$

- Find the deviations from the mean and square them

$$(x_1 - \overline{x})^2 = (78 - 83)^2 = (-5)^2 = 25$$

$$(x_2 - \overline{x})^2 = (82 - 83)^2 = (-1)^2 = 1$$

$$(x_3 - \overline{x})^2 = (89 - 83)^2 = (6)^2 = 36$$

- Sum of squared deviations

$$\text{SSD} = 25 + 1 + 36$$

- Calculate variance of sample

$$s^2 = \frac{\text{SSD}}{n - 1} = \frac{62}{3 - 1} = \frac{62}{2} = 31$$

- Calculate Sample Standard Deviation

$$s = \sqrt{s^2} = \sqrt{32} = 5.57$$

Step 5: Calculate the Test Statistic

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{83 - 85}{\frac{5.57}{\sqrt{3}}} = -\frac{3}{3.22} = -0.62$$

Step 6: Determine the Degrees of Freedom

$$df = n - 1 = 15 - 1 = 14$$

Step 7: Find Critical t-value

- For a two-tailed test at a significance level (α) of 0.05 and 2 degrees of freedom (df)

$$4.303$$

Step 8: Compare the t-Value to the Critical t-Value

- If the absolute value of the test statistic is greater than the critical t-value, reject the null hypothesis.
- If the absolute value of the test statistic is less than the critical t-value, fail to reject the null hypothesis.

Step 8: Find the p-Value

$$t = 4.303 \text{ and } df = 3$$

$$\text{p-value} = 0.58$$

```
from scipy import stats
rvs = stats.uniform.rvs(size=50)
stats.ttest_1samp(rvs, popmean=0.5)
```

21.3.2. Independent

Compares the means of two independent samples.

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}}$$

- $\overline{x}_1$ and $\overline{x}_2$: sample means
- s_1^2 and s_2^2: sample variances
- s_1 and n_2: sample sizes

```
from scipy import stats
rvs1 = stats.norm.rvs(loc=5, scale=10, size=500)
rvs2 = stats.norm.rvs(loc=5, scale=10, size=500)
stats.ttest_ind(rvs1, rvs2)
```

21.3.3. Paired

$$t = \frac{\overline{D}}{\frac{s_D}{\sqrt{n}}}$$

- $\overline{D}$: mean of the differences between paired observations
- s_D: standard deviation of the differences
- n: number of pairs

21.4. Chi-square tests

21.4.1. Goodness of Fit Test

Compares an **observed** categorical distribution to a **theoretical** categorical distribution.

$$\mathrm{X}^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

- o_i: observed frequency in category i
- e_i: expected frequency in category i

```
from scipy import stats
f_obs = np.array([43, 52, 54, 40])
f_exp = np.array([47, 47, 47, 47])
stats.chisquare(f_obs=f_obs, f_exp=f_exp)
```

21.4.2. Test of independence

Compares two **observed** categorical distributions.

$$\mathrm{X}^2 = \sum_{i=1}^{k} \frac{\left(o_{ij} - e_{ij}\right)^2}{e_{ij}}$$

- o_{ij}: observed frequency in cell (i, j)
- e_{ij}: expected frequency in category (i, j) calculated as:

$$E_{ij} = \frac{R_i \times C_j}{N}$$

- R_i: Row total for row i
- C_j: Column total for column j
- N: Total number of observations

21.5. ANOVA (Analysis of Variance)

21.5.1. One-way

Compares the means of three or more groups based on one independent variable

- H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
- H_1: At least one μ_i differs from the others

Step 1: Calculate Between-Group Variation (SS_{between})

$$SS_{\text{between}} = \sum_{i=1}^{k} n_i \left(\overline{X}_i - \overline{X}_{\text{overall}}\right)^2$$

- n_1: Number of observations in group i
- $\overline{X}_i$: Mean of group i
- $\overline{X}_{\text{overall}}$: Overall mean of all groups

Step 2: Calculate Within-Group Variation (SS_{within})

$$SS_{\text{within}} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}_i\right)^2$$

- X_{ij}: Observation j in group i

Step 3: Calculate Total Variation (SS_{total})

$$SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}}$$

Step 4: Calculate Mean Squares

- Mean Square Between (MS_{between})

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{k - 1}$$

- Mean Square Within (MS_{within})

$$MS_{\text{within}} = \frac{SS_{\text{within}}}{N - k}$$

- N: total number of observations
- k: number of groups

Step 5: Calculate the F-statistic

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

Step 6: Decision Rule

- Compare the F-statistic to the critical value from the F-distribution table (based on chosen significance level α).
- Alternatively, compare the p-value to the significance level α.
- Reject H_0 if the F-statistic is greater than the critical value or if the p-value is less than α, indicating that at least one group mean is significantly different.

21.5.2. Two-way

22. Regression Analysis:

22.1. Simple linear regression

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

- y: dependent variable
- x: independent variables
- β_0: intercept (value of y when $x = 0$)
- β_1: slope (change in y for a one-unit change in x)
- ε: error term (difference between the actual data points and the predicted values)

Data

x (Hours Studied)	y (Test Score)
1	2
2	3
3	5
4	4
5	6

Step 1: Calculate Means

$$\overline{x} = 3$$

$$\overline{y} = 4$$

Step 2: Calculating Slope β_1

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

- The numerator

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 9$$

- The denominator

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = 10$$

- The slope β_1 is

$$\beta_1 = \frac{9}{10} = 0.9$$

Step 3: Calculate Intercept β_0

$$\beta_1 = \overline{y} - \beta_1\overline{x} = 1.3$$

Step 4: Calculate p-Value

- Calculate Standard Error of the Slope (SE_{β_1})

$$SE_{\beta_1} = \sqrt{\frac{\sum_{n=1}^{n}(y_i - \hat{y}_i)^2}{(n-2) \cdot \sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

- Calculate the Residual Sum of Squares (RSS)

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Calculate the t-statistic for the Slope

$$t = \frac{\beta_1}{SE_{\beta_1}}$$

- Determine Degrees of Freedom

$$\mathrm{df} = n - 2$$

- Look up the p-value corresponding to t with 3 degrees of freedom

Step 5: Calculate R^2 (R^2_{adj})

$$R^2 = \frac{\mathrm{SS}_{\text{reg}}}{\mathrm{SS}_{\text{total}}}$$

SS_{reg} (Regression Sum of Squares): sum of the squared differences between the predicted $\hat{y}$ values and the mean of the observed y values.

$$\text{SS}_{\text{reg}} = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2$$

SS_{total} (Total Sum of Squares): sum of the squared differences between the observed y values and the mean of the observed y values.

$$\text{SS}_{\text{total}} = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

Adjusting for the numer of independent variables

$$R^2_{\text{adj}} = 1 - \left(\frac{(1 - R^2)(n - 1)}{n - k - 1} \right)$$

- n: number of observations
- k: number of independent variables

22.2. Multiple regression

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1 + ... + \beta_n x_n + \varepsilon$$

Matrix form:

$$Y = X\beta + \varepsilon$$

Estimating Coefficients (OLS):

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Where:

- $\hat{\beta}$: vector of estimated coefficients

22.3. Logistic Regression

Binary classification (1 or 0, true or false, yes or no) based on one or more predictor variables

Sigmoid function:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Where:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$$

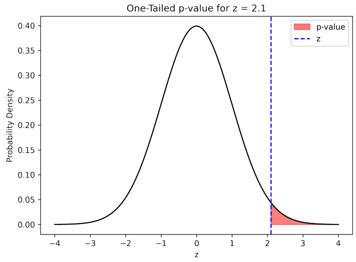
22.4. Model diagnostics

22.4.1. p-Values

Probability of obtaining results at least as extreme as the observed results

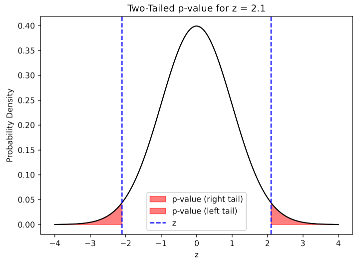
1. One-Tailed

$$p = P(Z > z_{\text{observed}})$$



2. Two-Tailed

$$p = 2 \cdot P(Z > |z_{\text{observed}}|)$$




```
z = 2.1
df = 3
scipy.stats.norm.sf(abs(z), df=df)
```

22.4.2. F-Statistic

$$F = \frac{\text{MSR}}{\text{MSE}}$$

Where:

- Mean Square Regression (MSR):

$$\text{MSR} = \frac{\text{SSR}}{\text{df}_{\text{regression}}}$$

- $\text{df}_{\text{regression}} = p$

Where:

- p: Number of independent
- Mean Square Error (MSE):

$$\text{MSE} = \frac{\text{SSE}}{\text{df}_{\text{error}}}$$

- $\text{df}_{\text{error}} = n - p - 1$

Where:

- p: Number of independent
- n: Number of observations
- 1: Constant (for intercept)

22.4.3. Confidence Intervals (CI)

Range within which we can be confident that the true value (population parameter) lies, based on the sample data

1. Known population standard deviation (σ):

$$CI = \overline{x} \pm z\frac{\sigma}{\sqrt{n}}$$

Where:

- $\overline{x}$: sample mean
- z: z-score corresponding to the desired confidence level
- σ: population standard deviation
- n: sample size

2. Unknown population standard deviation (σ):

$$CI = \overline{x} \pm t\frac{s}{\sqrt{n}}$$

Where:

- $\overline{x}$: sample mean
- t: critical value from t-distribution
- s: sample standard deviation
- n: sample size

23. Correlation

23.1. Pearson

Linear relationships

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \times (y_i - \overline{y})^2}}$$

Interpretation:

- 1: perfect positive linear relationship
- −1: perfect negative linear relationship
- 0: no linear relationship

$$X : [1, 2, 3, 4, 5]$$
$$Y : [2, 4, 6, 8, 10]$$

Step 1: Calculate the Means of X and Y

$$\overline{X} = 3$$
$$\overline{Y} = 6$$

Step 2: Calculate the Differences from the Mean

23.2. Spearman's Rank

Non-linear relationships

$$\rho = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

Where

- ρ: Spearman rank correlation coefficient
- d_i: difference between the ranks of corresponding values
- n: number of observations

Rank: position of a value within a data set when the values are ordered in ascending order

Observation i	X	Rank X	Y	Rank Y	d_i = Rank X - Rank Y	d_i^2
1	3	3	8	5	−2	4
2	1	1	6	3	−2	4
3	4	4	7	4	0	0
4	2	2	4	1	1	1
5	5	5	5	2	3	9

Interpretation

- $\rho = 1$: Perfect positive correlation
- $\rho = -1$: Perfect negative correlation
- $\rho = 0$: No correlation

24. Non-Parametric Statistics

24.1. Mann-Whitney U

Determine whether there is a significant difference between the distributions of two independent samples (used as an alternative to the independent samples t-test when the assumptions of normality are not met)

24.2. Wilcoxon Signed-Rank

Compare two paired samples or to assess whether the median of a single sample is different from a specified value (used as a non-parametric alternative to the paired t-test when the data does not meet the assumptions of normality)

24.3. Kolmogorov-Smirnov

Determine if a sample is drawn from a population with a specific distribution.

It compares the empirical distribution function (EDF) of the sample with the cumulative distribution function (CDF) of the reference distribution.

24.4. Kruskal-Wallis

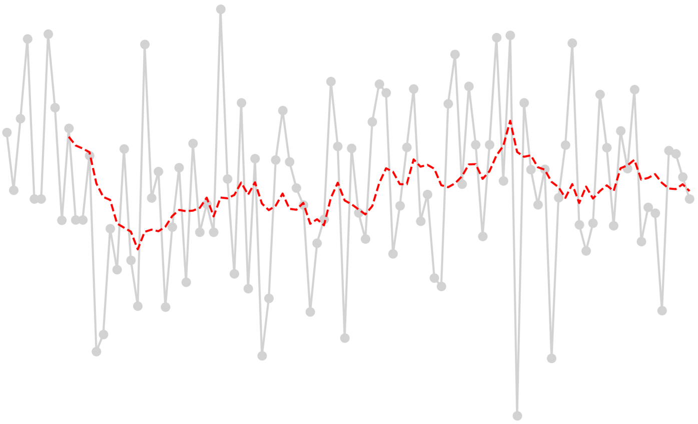
Determine if there are significant differences between the medians of three or more independent groups (extension of the Mann-Whitney U test, which is used for comparing two groups)

25. Time Series

25.1. SMA (Simple Moving Averages)

Creating a series of averages of different subsets (i.e., window) of the full data set

Minimize impact of short-term fluctuations



3-Day Simple Moving Average for the 7 Day Time Series

Data

- Day 1: $10
- Day 2: $12
- Day 3: $14
- Day 4: $16
- Day 5: $18
- Day 6: $20
- Day 7: $22

$$\text{SMA for Day 3} = \frac{\text{Day 1} + \text{Day 2} + \text{Day 3}}{3}$$

$$\text{SMA for Day 4} = \frac{\text{Day 2} + \text{Day 3} + \text{Day 4}}{3}$$

$$\text{SMA for Day 5} = \frac{\text{Day 3} + \text{Day 4} + \text{Day 5}}{3}$$

$$\text{SMA for Day 6} = \frac{\text{Day 4} + \text{Day 5} + \text{Day 6}}{3}$$

$$\text{SMA for Day 7} = \frac{\text{Day 5} + \text{Day 6} + \text{Day 7}}{3}$$

```
pl.col('X').rolling_mean(window_size)
```

25.2. WMA (Weighted Moving Average)

Each data point in the window is assigned a specific weight (usually decrease linearly)

$$\text{WMA} = \frac{\sum_{i=1}^{n}(x_i w_i)}{\sum_{i=1}^{n} w_i}$$

Data

- Day 1: $10
- Day 2: $12
- Day 3: $14
- Day 4: $13
- Day 5: $15

Weights

- 1st most recent: 3
- 2nd most recent: 2
- 3rd most recent: 1

$$\text{WMA for Day 3} = \frac{(14 \times 3) + (12 \times 2) + (10 \times 1)}{3 + 2 + 1} = 12.67$$

$$\text{WMA for Day 4} = \frac{(13 \times 3) + (14 \times 2) + (12 \times 1)}{3 + 2 + 1} = 13.17$$

$$\text{WMA for Day 5} = \frac{(15 \times 3) + (13 \times 2) + (14 \times 1)}{3 + 2 + 1} = 14.7$$

25.3. Exponential Smoothing

Gives more weight to recent data points (more responsive to new information)

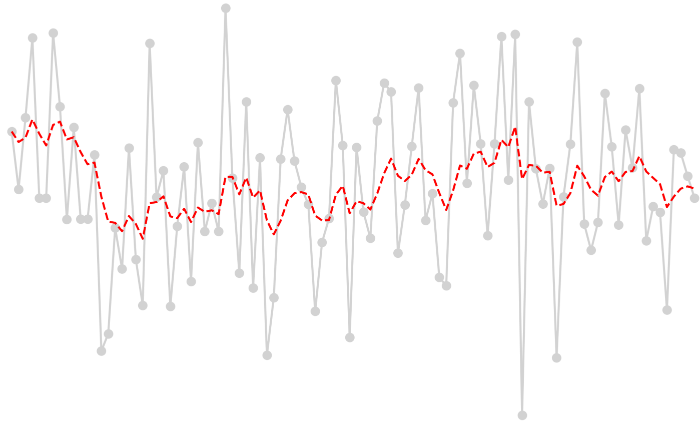
$$\text{EMA}_t = \alpha X_t + (1 - \alpha)\text{EMA}_{t-1}$$

Where

- α: smoothing factor

$$\alpha = \frac{2}{N + 1}$$

- N: Period (i.e., window size) of the EMA



Data

- Day 1: $10
- Day 2: $12
- Day 3: $14
- Day 4: $13
- Day 5: $15

Calculate Smoothing Factor

$$\alpha = \frac{2}{3+1} = 0.5$$

Calculate SMA for First Value

$$\text{SMA for Day 3} = \frac{10+12+14}{3}$$

Calculate EMA

$$\text{EMA for Day 4} = (0.5 \times 13) + (12 \times 0.5) = 12.5$$

$$\text{EMA for Day 5} = (0.5 \times 15) + (12.5 \times 0.5) = 13.75$$

```
pl.col("X").ewm_mean(span=window_size, adjust=False)
```

25.4. Seasonal Decomposition

- Level: baseline value around which the time series fluctuates
- Trend: ong-term progression or direction of the time series
- Seasonality: regular, repeating patterns or cycles in the time series
- Noise: random fluctuations or irregular variations (cannot be explained by the trend, seasonality, or other components)

Additive Decomposition

When the magnitude of seasonal fluctuations and trend does not change over time

$$Y_t = L_t + T_t + S_t + N_t$$

Where

- L_t: Level at time t
- T_t: Trent at time t
- S_t: Seasonal component at time t
- N_t: Noise (residuals) at time t

Multiplicative Decomposition

When the magnitude of seasonal fluctuations and trend change proportionally over time

$$Y_t = L_t \times T_t \times S_t \times N_t$$

Where

- L_t: Level at time t
- T_t: Trent at time t
- S_t: Seasonal component at time t
- N_t: Noise (residuals) at time t

25.5. ARMA (AutoRegressive Moving Average)

25.6. ARIMA (AutoRegressive Integrated Moving Average)

Industrial Engineering & Operations Research

26. Causes of variation

26.1. Common

Common Cause Variation: Normal, expected variation inherent to the process; predictable and stable.

- Temperature
- Humidity
- Haterial properties

26.2. Special

Special Cause Variation: Abnormal, unexpected variation due to specific causes; unpredictable and requires immediate action.

External disruptions

- Machine malfunction
- Faulty raw materials
- Staff shortages
- Weather events

27. Design of Experiments (DOE):

28. Control Charts:

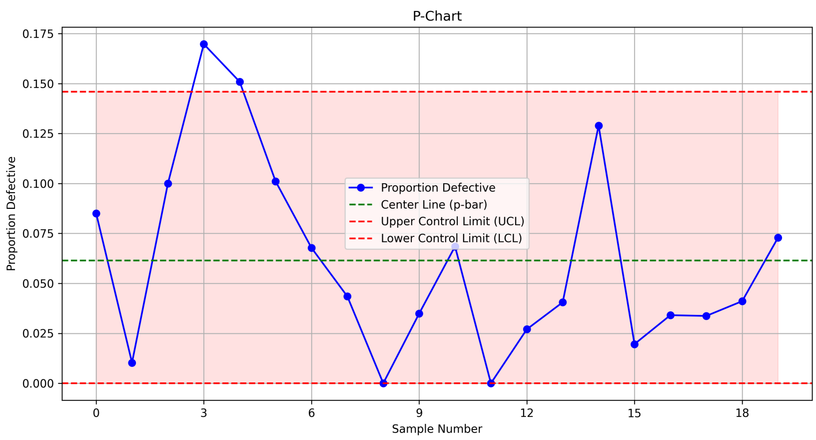
28.1. P-charts (Proportion)

Proportion of defective items

$$\hat{p} = \frac{D}{n}$$

Where:

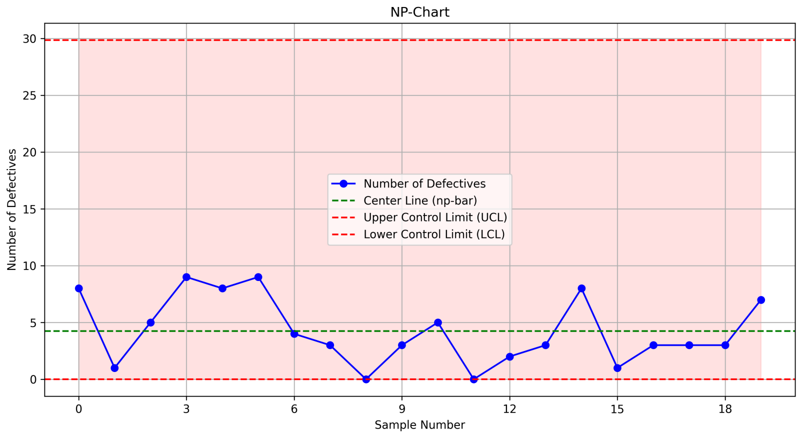
- D: number of defective items
- n: sample size



28.2. NP-charts (Number Proportion)

Number of defective items (constant sample size)

- N: count of defective items in each sample



28.3. C-charts (Count)

Count of defects (fixed unit size)

Number of defects observed in each sample or inspection unit

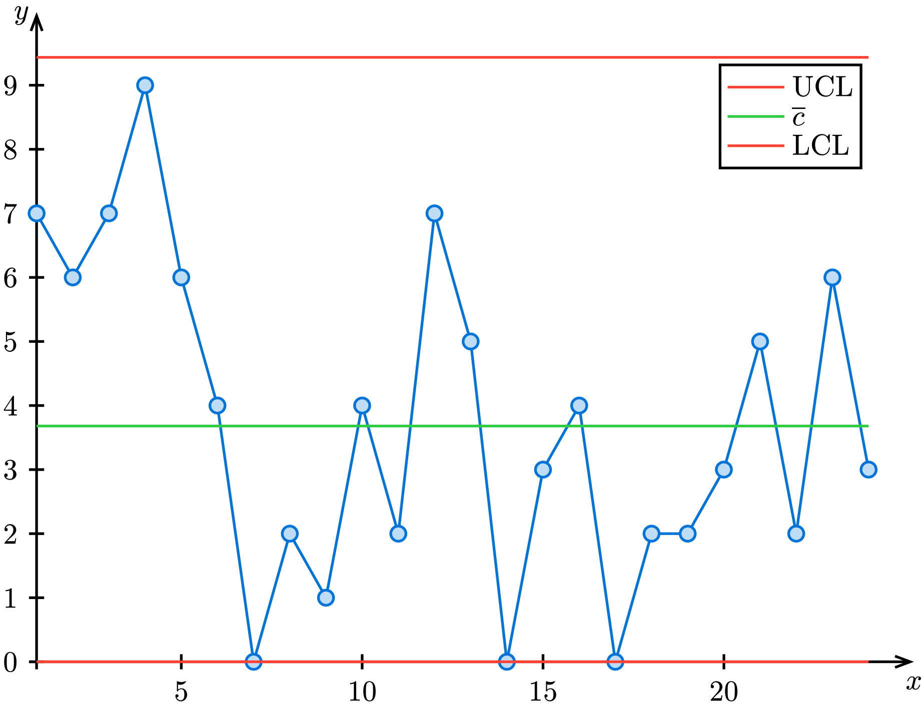
$$\overline{c} = \frac{\sum_{i=1}^{k} c_i}{k}$$

$$\mathrm{UCL}_c = \overline{c} + 3\sqrt{\overline{c}}$$

$$\mathrm{LCL}_c = \overline{c} - 3\sqrt{\overline{c}}$$

Where:

- c: number of defects
- k: number of samples



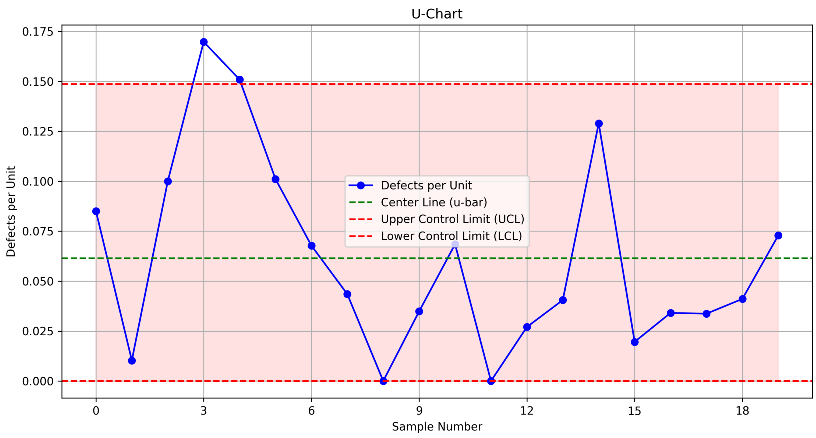
28.4. U-charts (Unit)

Defects per unit (variable unit size)

$$\hat{u} = \frac{C}{n}$$

Where:

- C: number of defects
- n: size of unit



28.5. $\overline{X}$-chart

28.6. R-chart

29. Process Capability Analysis

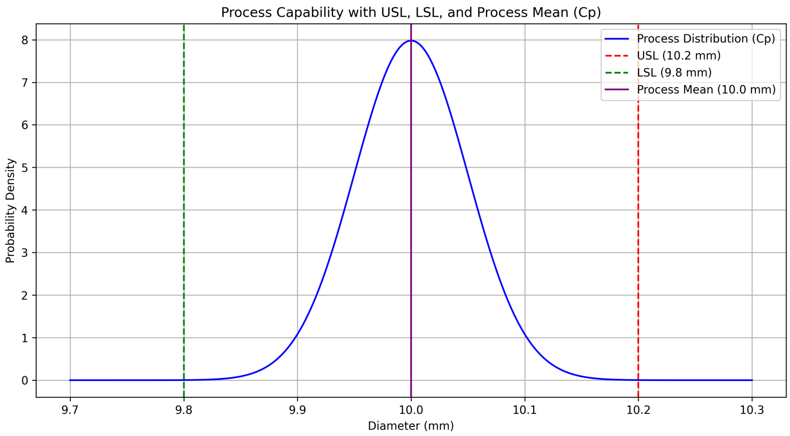
29.1. C_p (Process Capability Index)

Measure how well a process can produce outputs within specified limits

$$C_p = \frac{\text{USL} - \text{LSL}}{6\sigma}$$

- $C_p > 1$: The process variation is smaller than the specification range (good capability).
- $C_p = 1$: The process variation matches the specification range (barely acceptable).
- $C_p < 1$: The process variation exceeds the specification range (poor capability).

Assumption: Process is **centered** within the specification limits



Suppose a company manufactures metal rods, and the specification limits for the diameter of the rods are:

- Upper Specification Limit (**USL**): 10.2 mm
- Lower Specification Limit (**LSL**): 9.8 mm

The process has a standard deviation **0.05** of 0.05 mm.

Step 1: Determine the Specification Width

The specification width is the difference between the USL and LSL.

$$\text{Specification Width} = \text{USL} - \text{LSL} = 10.2 \text{ mm} - 9.8 \text{ mm}$$

Step 2: Calculate the Process Capability Index C_p

The formula for C_p is:

$$C_p = \frac{\text{Specification Width}}{60.05} = \frac{\text{USL} - \text{LSL}}{60.05}$$

Substitute the values:

$$C_p = \frac{0.4 \text{ mm}}{6 \times 0.05 \text{ mm}} = \frac{0.4 \text{ mm}}{0.3 \text{ mm}} = 1.33$$

Interpretation:

C_p = 1.33 means the process spread (6 0.05) fits 1.33 times within the tolerance range (the distance between the Upper Specification Limit and Lower Specification Limit.

- C_p = 1.00: Process variation fits exactly within the specification limits. 99.73% of the output will be within specifications **if the process is centered** (3 sigma process).
- $C_p > 1.00$: Process variation is narrower than the specification limits. The higher the C_p, the more capable the process is, meaning it can produce parts within the tolerance more consistently.
- $C_p < 1$: Process variation is wider than the specification limits. Significant portion of the output will fall outside the specification limits.

Limitations:

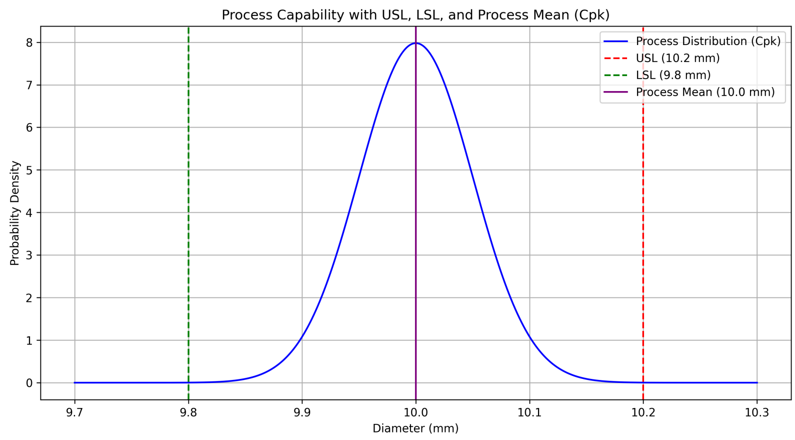
Since **C_p does not account for the centering of the process**, it may give a false sense of security if the process mean is off-center (**see C_{pk}**).

Use:

When you are interested in understanding the **potential capability** of a process under ideal conditions, typically in a short-term study where the process is stable and controlled.

29.2. C_{pk} (Process Capability Index with Centering)

$$C_{pk} = \text{Min}\left(\frac{\text{USL} - \overline{x}}{3\sigma}, \frac{\overline{x} - \text{USL}}{3\sigma}\right)$$



Suppose a company manufactures metal rods, and the specification limits for the diameter of the rods are:

- Upper Specification Limit (**USL**): 10.2 mm
- Lower Specification Limit (**LSL**): 9.8 mm

The process has:

- A standard deviation σ of 0.05 mm.

- A process mean μ of 10.1 mm.

Step 1: Calculate the distance from the mean to the USL:

$$\frac{\text{USL} - \mu}{3\sigma} = \frac{10.2 \text{ mm} - 10.1 \text{ mm}}{3 \times 0.05 \text{ mm}} = \frac{0.1 \text{ mm}}{0.15 \text{ mm}} = 0.67$$

Step 2: Calculate the distance from the mean to the LSL:

$$\frac{\mu - \text{LSL}}{3\sigma} = \frac{10.1 \text{ mm} - 9.8 \text{ mm}}{3 \times 0.05 \text{ mm}} = \frac{0.3 \text{ mm}}{0.15 \text{ mm}} = 2.00$$

Step 3: Determine C_{pk}:

$$C_{pk} = \min(0.67, 2.00) = 0.67$$

Interpretation:

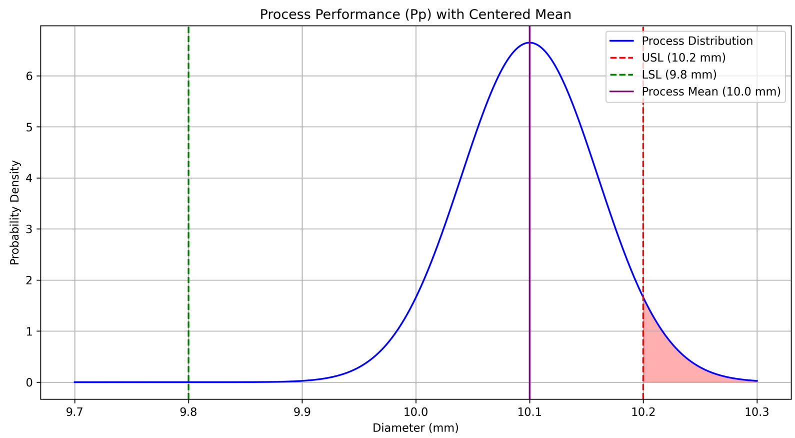
C_p = 1.33 means the process spread (6 σ) fits 1.33 times within the tolerance range (the distance between the Upper Specification Limit and Lower Specification Limit.

- C_p = 1.00: Process mean is exactly at the midpoint of the specification limits, and the process variation fits exactly within these limits. 99.73% of the output will be within specifications, indicating a capable process (3 sigma process).
- $C_p > 1.00$: The higher the C_{pk}, the more capable and stable the process is, meaning it can consistently produce parts within tolerance with minimal risk of defects.
- $C_p < 1$: The process mean is off-center or the variation is wider than the specification limits, or both. A significant portion of the output may fall outside the specification limits.

29.3. C_{pm} (Taguchi Capability Index)

29.4. P_p (Process Performance Index)

$$P_p = \frac{\text{USL} - \text{LSL}}{6\sigma_{\text{overall}}}$$



Suppose a company manufactures metal rods, and the specification limits for the diameter of the rods are:

- Upper Specification Limit (**USL**): 10.2 mm
- Lower Specification Limit (**LSL**): 9.8 mm
- Overall standard deviation (σ_{overall}): 0.06 mm

Step 1: Determine the Specification Width

The specification width is the difference between the USL and LSL.

$$\text{Specification Width} = \text{USL} - \text{LSL} = 10.2\text{ mm} - 9.8\text{ mm} = 0.4$$

Step 2: Calculate the Process Performance Index P_p

The formula for P_p is:

$$P_p = \frac{\text{Specification Width}}{6\sigma_{\text{overall}}} = \frac{\text{USL} - \text{LSL}}{6\sigma_{\text{overall}}}$$

Substitute the values:

$$P_p = \frac{0.4\text{ mm}}{6 \times 0.06\text{ mm}} = \frac{0.4\text{ mm}}{0.36\text{ mm}} = 1.11$$

Interpretation:

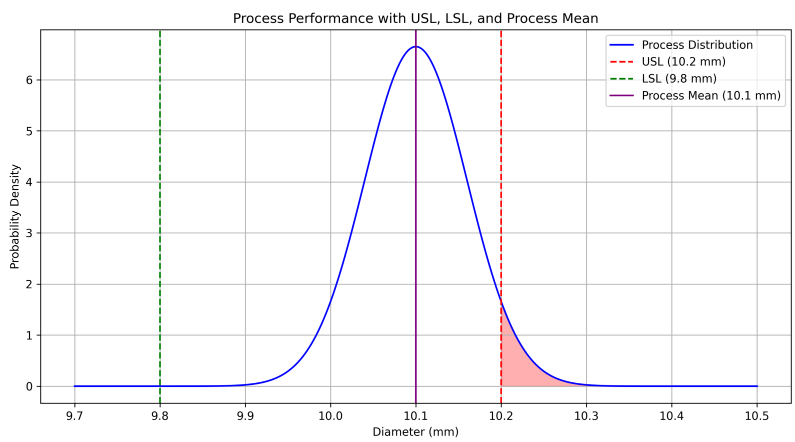
- A P_p of 1.11 indicates that the process performance, considering all sources of variation over time, is capable but less so than the potential capability indicated by C_p. The value being slightly above 1 suggests that the process can generally produce rods within specifications, but there might be more variability in the process compared to the short-term capability measured by C_p
- Decrease from the C_p value (1.33 to 1.11) reflects the impact of additional variability when evaluating the process over a longer time or under different conditions.

Use:

When you need to evaluate the **actual performance** of a process over a longer period, considering all sources of variation, including shifts, drifts, and other long-term factors.

29.5. P_{pk} (Process Performance Index with Centering)

$$P_{pk} = \text{Min}\left(\frac{\text{USL} - \mu_{\text{overall}}}{3\sigma}, \frac{\mu_{\text{overall}} - \text{USL}}{3\sigma}\right)$$



Suppose a company manufactures metal rods, and the specification limits for the diameter of the rods are:

- Upper Specification Limit (**USL**): 10.2 mm
- Lower Specification Limit (**LSL**): 9.8 mm
- Overall standard deviation (σ_{overall}): 0.06 mm
- Overall process mean (μ_{overall}): 10.1 mm

Step 1: Calculate the Distance from the Process Mean to the Specification Limits

Calculate the distance from the overall process mean to both the USL and LSL:

$$\text{USL} - \mu_{\text{overall}} = 10.2 \text{ mm} - 10.1 \text{ mm} = 0.1 \text{ mm}$$

$$\mu_{\text{overall}} - \text{LSL} = 10.1 \text{ mm} - 9.8 \text{ mm} = 0.3 \text{ mm}$$

Step 2: Calculate the Process Performance Index P_k

The formula for P_{pk} is:

$$P_{pk} = \text{Min}\left(\frac{\text{USL} - \mu_{\text{overall}}}{3\sigma}, \frac{\mu_{\text{overall}} - \text{USL}}{3\sigma}\right)$$

Substitute the values:

$$P_{pk} = \text{Min}\left(\frac{0.1 \text{ mm}}{3 \times 0.06 \text{ mm}}, \frac{0.3 \text{ mm}}{3 \times 0.06 \text{ mm}}\right)$$

$$P_{pk} = \text{Min}\left(\frac{0.1 \text{ mm}}{0.18 \text{ mm}}, \frac{0.3 \text{ mm}}{0.18 \text{ mm}}\right)$$

$$P_{pk} = \min(0.56, 1.67) = 0.56$$

Interpretation:

- A P_p of 1.11 indicates that the process performance, considering all sources of variation over time, is capable but less so than the potential capability indicated by C_p. The value being slightly above 1 suggests that the process can generally produce rods within specifications, but there might be more variability in the process compared to the short-term capability measured by C_p
- Decrease from the C_p value (1.33 to 1.11) reflects the impact of additional variability when evaluating the process over a longer time or under different conditions.

Use:

When you need to evaluate the **actual performance** of a process over a longer period, considering all sources of variation, including shifts, drifts, and other long-term factors.

30. Inventory Management

30.1. Newsvendor

determine the optimal order quantity Q^* that minimizes the total expected cost or maximizes the expected profit, based on the trade-off between the overage and underage costs

Assumptions

- Products are separable

- Planning is done for a single period
- Demand is random
- Deliveries are made in advance of demand
- Costs of overage or underage are linear

1. Parameters
    - P: Sale Price
    - C: Purchase Cost
    - S: Unsold Value
    - μ: Mean Demand
    - σ: Standard Deviation Demand

2. Calculate Underage and Overage Costs:
    - Underage Cost (C_u): Profit lost for each unit of demand not met

    $$C_u = P - C$$

    - Overage Cost (C_o): This is the cost of holding an unsold newspaper.

    $$C_o = C - S$$

3. Calculate Critical Ratio (CR)

$$CR = \frac{C_u}{C_u - C_o}$$

4. Find z-score: Find the number of standard deviations away from the mean corresponding to the critical ratio:

$$z^* = \Phi^{-1}(CR)$$

Where:
- Φ^{-1}: Inverse of the CDF of the standard normal distribution (PPF)

5. Calculate Optimal Order Quantity (Q^*)

$$Q^* = \mu + z^* \sigma$$

Where:
- z^*: z-score corresponding to the critical ratio CR from the standard normal distribution

Consider a newsvendor selling newspapers:

1. Parameters
    - Sale Price (P): \$3 (per unit)
    - Purchase Cost (C): \$1 (per unit)
    - Unsold Value (S): \$0 (per unit)
    - Mean Demand (μ): 100 (units)
    - Standard Deviation Demand (σ): 20 (units)

2. Calculate Underage and Overage Costs:
    - Underage Cost (C_u): Profit lost for each unit of demand not met

    $$C_u = P - C = 3 - 1 = 2$$

- Overage Cost (C_o): This is the cost of holding an unsold newspaper.

$$C_o = C - S = 1 - 0 = 1$$

3. Calculate Critical Ratio (CR)

$$CR = \frac{C_u}{C_u - C_o} = \frac{2}{2+1} = \frac{2}{3} = 0.67$$

4. Find z-score: Find the number of standard deviations away from the mean corresponding to the critical ratio:

$$z^* = \Phi^{-1}(CR) = 0.44$$

5. Calculate Optimal Order Quantity (Q^*)

$$Q^* = \mu + z^* \sigma = 100 + 0.44 \cdot 20 = 108.8$$

```
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

# Parameters for the newsvendor example
selling_price = 3  # Selling price per newspaper
purchase_cost = 1   # Purchase cost per newspaper
unsold_value = 0    # Value of unsold newspapers
mu = 100 # Mean of demand
sigma = 20 # Standard deviation of demand

# Calculate underage and overage costs
C_u = selling_price - purchase_cost  # Underage cost
C_o = purchase_cost - unsold_value    # Overage cost

# Calculate the critical ratio
CR = C_u / (C_u + C_o)

# Assume a normal distribution for demand
mu = 50    # Mean demand
sigma = 10 # Standard deviation of demand

# Calculate the optimal order quantity (Q*)
z_star = stats.norm.ppf(CR)  # z-score corresponding to the critical ratio
Q_star = mu + z_star * sigma  # Optimal order quantity
```

30.2. ABC Analysis

categorizes inventory items into three groups (A, B, and C) based on their importance

- A: Top 70-80% of the total annual consumption value
- B: Next 15-25% of the total annual consumption value
- C: Remaining 5-10% of the total annual consumption value

Item	Usage Quantity (Annual)	Unit Cost	Conumption Value (Annual)
I_1	50	$100	$5000
I_2	150	$20	$3000
I_3	300	$10	$3000
I_4	400	$5	$2000
I_5	500	$1	$500

Step 1: Calculate Annual Consumption Values

Step 2: Sort Items by Annual Consumption Value (Descending)

Step 3: Calculate Total Annual Consumption Value

$$\text{Total} = \$5,000(I_1) + \$3,000(I_2) + \$3,000(I_3) + \$2,000(I_4) + \$500(I_5) = \$13500$$

Step 4: Calculate Cumulative Consumption Value Percentages

- $I_1 : \frac{5000}{13500} \times 100\% = 37.04\%$
- $I_2 : \frac{3000}{13500} \times 100\% = 22.22\%$
- $I_3 : \frac{3000}{13500} \times 100\% = 22.22\%$
- $I_4 : \frac{2000}{13500} \times 100\% = 14.81\%$
- $I_5 : \frac{500}{13500} \times 100\% = 3.7\%$

Step 5: Cumulative Percentages

- I_1: 37.04%
- $I_1 + I_2$: 59.26%
- $I_1 + I_2 + I_3$: 81.48%
- $I_1 + I_2 + I_3 + I_4$: 96.3%
- $I_1 + I_2 + I_3 + I_4 + I_5$: 100%

Step 6: Categorize Items

- A: I_1, I_2, I_3
- B: I_4
- C: I_5

30.3. Fill Rate

Percentage of customer demand that is satisfied from available inventory

$$F = \frac{U_f}{U_o} \times 100\%$$

Where:

- F: Fill Rate
- U_f: Number of units fulfilled
- U_o: Total number of units ordered

Customers order 100 units of a product and 90 units are fulfilled from stock:

$$F = \frac{90}{100} \times 100\% = 90\%$$

30.4. (OCT) Order Cycle Time

Measures the total time taken from when a customer places an order to when the order is delivered

$$\text{OCT} = T_{\text{order}} + T_{\text{processing}} + T_{\text{production}} + T_{\text{shipping}}$$

Where:

- T_{order}: Order Entry Time (time it takes to receive and log the order)
- $T_{\text{processing}}$: Order Processing Time (time to check inventory, verify details, and prepare for production or shipment)
- $T_{\text{production}}$: Production Time (time to manufacture or prepare the product)
- T_{shipping}: Shipping Time (time it takes to deliver the product from the warehouse to the customer)

30.5. ROP (Reorder Point)

The inventory level at which a new order should be placed to avoid stockouts

$$\text{ROP} = (\text{Average Demand per Period} \times \text{Lead Time})$$

Suppose your business sells 50 units per week, and the lead time for a new order is 2 weeks. Using the formula:

$$\text{ROP} = 50(\text{units per week}) \times 2(\text{weeks}) = 100 \text{ units}$$

This means that when your inventory level drops to 100 units, you should place a new order to avoid running out of stock.

30.6. XYZ Analysis

Categorization based on variability

- X: Low variability
- Y: Moderate variability
- Z: High variability

Coefficient of Variation:

$$CV = \frac{\sigma}{\mu}$$

- X Items: Low CV
  - CV < k_1, where k_1 is the threshold value indicating low variability
- Y Items: Moderate CV

- ‣ CV < k_2, where k_2 is the moderate value indicating low variability
- Z Items: High CV
    - ‣ CV < k_3, where k_3 is the threshold value indicating high variability

Step 1: Collect Historical D../ata (12 months)

- Product A: [100, 105, 98, 102, 101, 104, 103, 100, 99, 100, 101, 102]
- Product B: [150, 155, 145, 160, 140, 150, 155, 150, 165, 155, 150, 140]
- Product C: [200, 180, 220, 190, 210, 240, 180, 230, 220, 210, 250, 190]

Step 2: Calculate the Mean and Standard Deviation

- Product A:
    - $\mu = 101$
    - $\sigma = 2$
- Product B:
    - $\mu = 150$
    - $\sigma = 8$
- Product C:
    - $\mu = 210$
    - $\sigma = 25$

Step 3: Calculate the Coefficient of Variation (CV)

- $CV = \frac{2}{101} = 0.0198$
- $CV = \frac{8}{150} = 0.0533$
- $CV = \frac{25}{210} = 0.1190$

Step 4: Categorization

- X: Product A (Low variability)
- Y: Product B (Moderate variability)
- Z: Product C (High variability)

30.7. EOQ (Economic Order Quantity)

Optimal order quantity that **minimizes the total cost**, which includes both **holding costs** and **ordering costs**

$$\text{EOQ} = \sqrt{\frac{2DS}{H}}$$

Where:

- D: Demand per time period
- S: Ordering cost per order
- H: Holding cost per unit, per time period

A company sells widgets and wants to determine the optimal order quantity for inventory.

- The annual demand for widgets (D) is 12,000 units.
- The cost to place an order (S) is $50.

The holding cost per unit per year (H) is $2.

$$\text{EOQ} = \sqrt{\frac{2 \times 12000 \times 50}{2}} = 775$$

The company should order 775 widgets each time they place an order to minimize the total cost, which includes both ordering and holding costs

30.7.1. Perfect Order Rate

Measures the percentage of orders delivered to customers in full, on time, and without any damage

$$\text{Perfect Order Rate} = \frac{\text{Number of Perfect Orders}}{\text{Total Number of Orders}} \times 100\%$$

Where:

- **Number of Perfect Orders**: The number of orders that are delivered on time, complete, and undamaged.
- **Total Number of Orders**: The total number of orders fulfilled within a specific period.

Suppose you received 1,000 orders over a quarter, and 900 of those orders were delivered on time, complete, and without damage.

$$\text{Perfect Order Rate} = \frac{900}{1000} \times 100\% = 90\%$$

30.8. Safety Stock

Additional quantity of inventory kept on hand to protect against uncertainties in demand or supply. buffer to prevent stockouts due to unexpected variations in demand or delays in supply.

1. Constant Demand & Constant Lead Time

$$\text{SS} = Z \times \sigma_D$$

Where:

- Z: Z-score corresponding to the desired service level
- σ: standard deviation of demand

2. Variable Demand & Constant Lead Time

$$\text{SS} = Z \times \sigma_D \times \sqrt{L}$$

Where:

- Z: Z-score corresponding to the desired service level
- σ_D: Standard deviation of demand per unit of time
- L: Lead time

3. Variable Lead Time & Constant Demand

$$\text{SS} = Z \times \overline{D} \times \sigma_L$$

Where:

- Z: Z-score corresponding to the desired service level
- $\overline{D}$: Average demand
- σ_L: Standard deviation of lead time

4. Variable Demand & Variable Lead Time

$$\text{SS} = Z \times \sqrt{\left(\overline{D}^2 \times \sigma_L^2\right) + \left(L \times \sigma_D^2\right)}$$

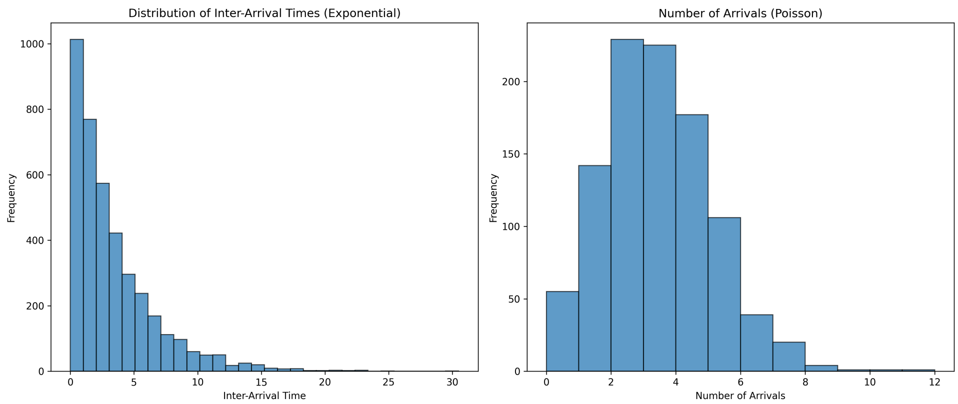
Where:

- $\overline{L}$: Average Lead Time (average time it takes to receive inventory after placing an order)
- σ_D^2: Demand Variance (variability in demand during the lead time)
- $\overline{D}$: Average Demand (mean quantity of demand per time period)
- σ_L^2: Lead Time Variance (variability in lead time)

31. Queuing Theory

31.1. M/M/1

- Arrival rate (λ): Average number of customers arriving per unit of time
- Service rate (μ): Average number of customers served per unit of time



1. Utilization (ρ)

Fraction of time the server is busy

$$\rho = \frac{\lambda}{\mu}$$

2. Average Number of Customers in the System (L)

Average number of customers (both waiting and being served) in the system

$$L = \frac{\rho}{1 - \rho}$$

3. Average Number of Customers in the Queue (L_q)

Average number of customers waiting in the queue

$$L_q = \frac{\rho^2}{1 - \rho}$$

4. Average Time a Customer Spends in the System (W)

Average time a customer spends in the system (from arrival until they are done being served)

$$W = \frac{1}{\mu - \lambda}$$

5. Average Waiting Time in the Queue (W_q)

Average time a customer spends just waiting in line before being served

$$W_q = \frac{\rho}{\mu - \lambda}$$

6. Probability that the System is Empty (P_0)

Probability that there are zero customers in the system (no one is being served and no one is waiting)

$$P_0 = 1 - \rho$$

7. Probability that n Customers are in the System (P_n)

Probability that there are n customers in the system (either waiting or being served)

$$P_n = (1 - \rho) \cdot \rho^n$$

8. Probability the Queue is Full (if the Queue has Limited Capacity) ($P_{n_{\max}}$)

Probability that the system is at full capacity

$$P_{n_{\max}} = (1 - \rho) \cdot p^{n_{\max}}$$

9. System Throughput

Rate at which customers are served and leave the system

$$\text{Throughput} = \lambda$$

10. Expected Time in Service (W_s)

Average time a customer spends actually being served (not including waiting time)

$$W_s = \frac{1}{\mu}$$

11. Idel Time ($1 - \rho$)

Fraction of time that the server is idle (i.e., not serving any customers)

$$\text{Idle Time} = 1 - \rho$$

12. Probability of Having to Wait in the Queue (P_w)

Probability that an arriving customer will have to wait before being served, i.e., that the server is busy when the customer arrives

$$P_w = \rho$$

13. Variance of the Number of Customers in the System ($\text{Var}(L)$)

Variance of the number of customers in the system

$$\text{Var}(L) = \frac{\rho}{(1 - \rho)^2}$$

A bank with a single teller

- **Arrival rate (λ)**: On average, 4 customers arrive every 10 minutes ($\lambda = 4$ custommers per 10 minutes)
- **Service Rate (μ)**: The teller can serve 6 customers every 10 minutes ($\mu = 6$ customers per 10 minutes)

1. **Utilization (ρ)**

$$\rho = \frac{\lambda}{\mu} = \frac{4}{6} = 0.67$$

The teller is busy 67% of the time. The remaining 33% of the time, the teller is idle, waiting for the next customer

2. Average Number of Customers in the System (L)

$$L = \frac{\rho}{1-\rho} = \frac{0.67}{1-0.67} = 2$$

On average, there are 2 customers in the coffee shop at any given time, either being served or waiting in line

3. Average Number of Customers in the Queue (L_q)

$$L_q = \frac{\rho^2}{1-\rho} = \frac{0.67^2}{1-0.67} = 1.33$$

On average, about 1.33 customers are waiting in line at any time

4. Average Time a Customer Spends in the System (W)

$$W = \frac{1}{\mu - \lambda} = \frac{1}{6} - 4 = 0.5$$

On average, a customer spends 5 minutes (0.5×10 minutes) in the shop (including both waiting in line and getting served)

5. Average Waiting Time in the Queue (W_q)

$$W_q = \frac{\rho}{\mu - \lambda} = \frac{0.67}{6-4} = 0.33$$

On average, a customer waits 3.3 minutes (0.33×10 minutes) in line before being served by the teller

6. Probability that the System is Empty (P_0)

$$P_0 = 1 - \rho = 1 - 0.67 = 0.33$$

There is a 33% chance that the coffee shop is empty, meaning there is no customer in the queue or being served

7. Probability that n Customers are in the System (P_n)

$$P_n = (1-\rho) \cdot \rho^n = (1-0.67) \cdot 0.67^2 = 0.148$$

There is a 14.8% chance that exactly 2 customers are either in line or being served

8. Probability the Queue is Full (if the Queue has Limited Capacity) ($P_{n_{\max}}$)

$$P_{n_{\max}} = (1 - \rho) \cdot p^{n_{\max}} = (1 - 0.67) \cdot 0.67^5 = 0.028$$

There is a 2.8% chance that the system is full, and no new customers can enter

9. System Throughput

$$\text{Throughput} = \lambda = 4$$

The coffee shop serves 4 customers every 10 minutes, on average

10. Expected Time in Service (W_s)

$$W_s = \frac{1}{\mu} = \frac{1}{6} = 0.167$$

On average, a customer spends 1.67 minutes being served by the teller

11. Idel Time ($1 - \rho$)

$$\text{Idle Time} = 1 - \rho = 1 - 0.67 = 0.33$$

The teller is idle 33% of the time

12. Probability of Having to Wait in the Queue (P_w)

$$P_w = \rho = 0.67$$

There is a 67% chance that a customer will have to wait when they arrive

13. Variance of the Number of Customers in the System ($\mathrm{Var}(L)$)

$$\mathrm{Var}(L) = \frac{\rho}{(1 - \rho)^2} = \frac{0.67}{(1 - 0.67)^2} = 6.12^2$$

The queue length varies significantly, with a variance of 6.12 customers

Costs

- **Cost per Waiting Customer per Unit Time (C_w)**: cost incurred per customer for each unit of time they spend waiting in the queue

$$\text{Total Waiting Cost} = L_q \times C_w \times \text{Unit Time}$$

- **Cost per Idle Server per Unit Time (C_s)**: cost incurred per unit time when the server is not serving customers

$$\text{Total Idle Cost} = (1 - \rho) \times C_s \times \text{Unit Time}$$

Total Cost

$$\text{Total Cost} = \text{Total Waiting Cost} + \text{Total Idle Cost}$$

- Arrival Rate (λ): 4 customers per 10 minutes
- Service Rate (μ): 6 customers per 10 minutes
- Cost per Waiting Customer per Hour (C_w): $10
- Cost per Idle Server per Hour (C_s): $20

- Operational Time: 1 hour

1. Utilization

$$\rho = \frac{\lambda}{\mu} = \frac{4}{6} = 0.67$$

2. Average Number of Customers in Queue (L_q)

$$L_q = \frac{\rho^2}{1-\rho} = \frac{0.67^2}{1-0.67} = 1.33 \text{ cusommers}$$

3. Total Waiting Cost

$$\text{Total Waiting Cost} = L_q \times C_w \times \text{Unit Time} = 1.33 \times 10 \times 1 = \$13.33$$

4. Idle Time

$$\text{Idle Time} = 1 - \rho = 1 - 0.67 = 0.33$$

5. Total Idle Cost

$$\text{Total Idle Cost} = (1-\rho) \times C_s \times \text{Unit Time} = 0.33 \times 20 \times 1 = \$6.67$$

6. Total Cost

$$\text{Total Cost} = \text{Total Waiting Cost} + \text{Total Idle Cost} = 13.33 + 6.67 = \$20$$

32. Network Optimization

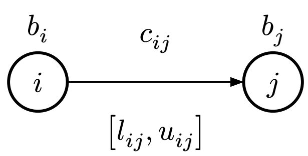
32.1. Shortest Path

32.2. Maximum Flow

32.3. Netwrok Flow Optimization

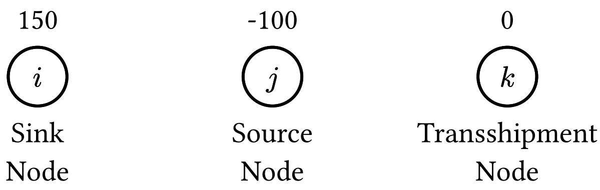
Types of Problems:

- Maximum Flow
- Minimum Cost Flow
- Multi-Commodity Flow



Node i

- Sink node: $b_i > 0$ — Has demand of b_i units
- Source node: $b_i < 0$ — Has supply of $-b_i$ units
- Transsipment node: $b_i = 0$ — Neither supply or demand



Units shipped from node i to node j:

$$x_{ij}$$

Minimize:

$$Z = \sum_{i,j} c_{ij} x_{ij}$$

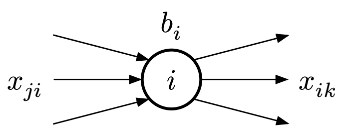
s.t.

$$\sum_k x_{ki} - \sum_l x_{il} = \text{ or } \leq \text{ or } \geq b_1$$

$$l_{ij} \leq x_{ij} \leq u_{ij}$$

Where:

- c_{ij}: Unit cost of flow from node i to node j
- b_i: Demand on node i
- l_{ij}: flow lower bound from i to j
- u_{ij}: Flow upper bound from i to j
- $\sum_k x_{ki}$: Inflow to i
- $\sum_l x_{il}$: Outflow from j



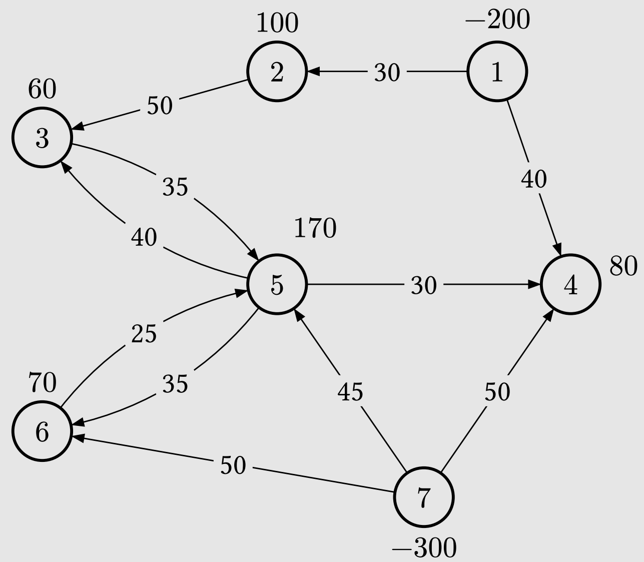
If:

- Total Supply $=$ Total Demand — [Inflow to i] - [Outflow from i] $= b_i$
- Total Supply $>$ Total Demand — [Inflow to i] - [Outflow from i] $\geq b_i$
- Total Supply $<$ Total Demand — [Inflow to i] - [Outflow from i] $\leq b_i$

Important:

- One decision variable x_{ij} for each edge (i, j)

- One flow balancing constraint for each node i



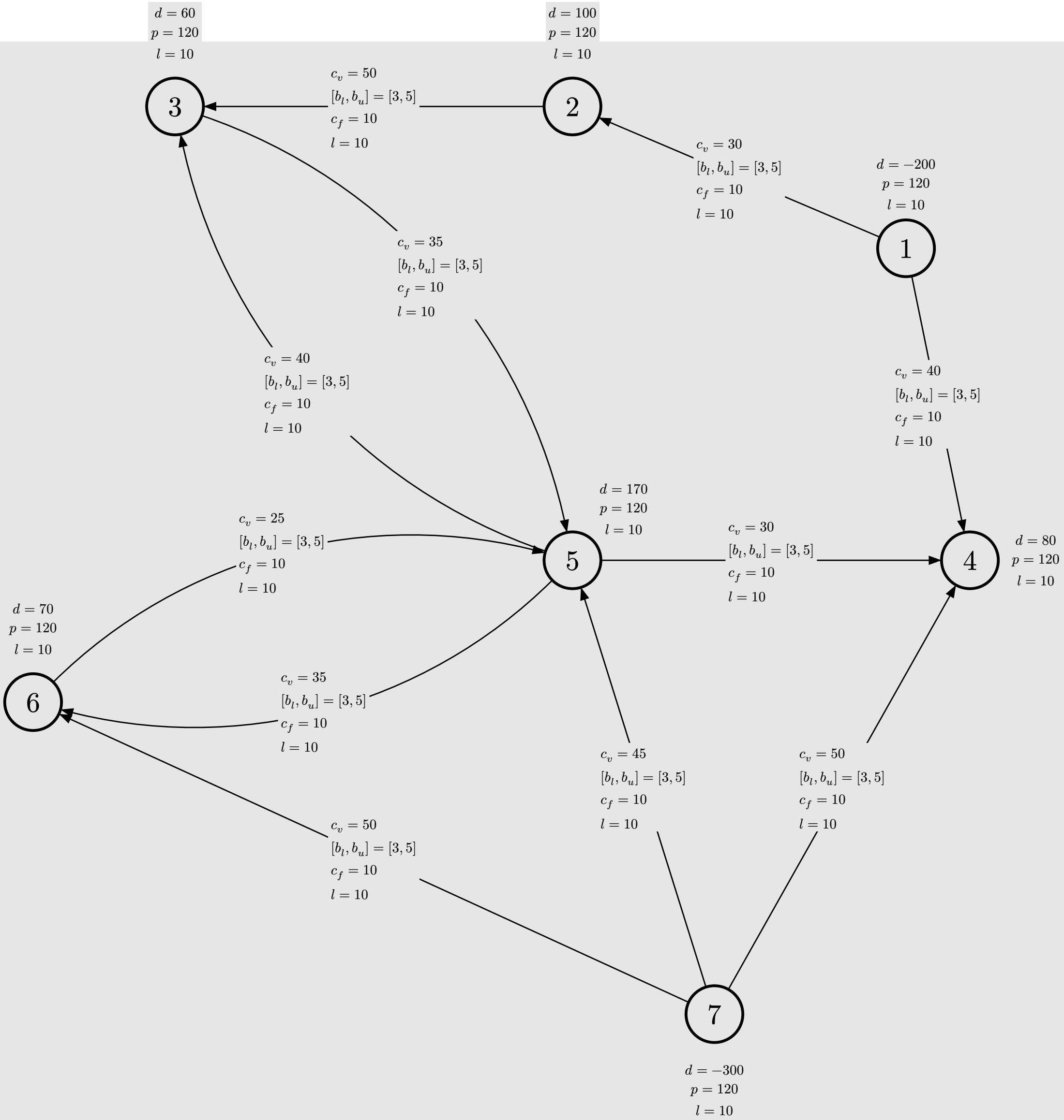
Minimize

$$Z = 30x_{12} + 40x_{14} + 50x_{23} + 35x_{35} + 40x_{53}$$
$$+30x_{54} + 35x_{56} + 25x_{65} + 50x_{74} + 45x_{75} + 50x_{76}$$

s.t.

$$x_{12} + x_{14} \leq 200 \qquad \text{(Node 1)}$$
$$x_{12} + x_{23} \geq 100 \qquad \text{(Node 2)}$$
$$x_{23} + x_{53} - x_{35} \geq 60 \qquad \text{(Node 3)}$$
$$x_{14} + x_{54} + x_{74} \geq 80 \qquad \text{(Node 4)}$$
$$x_{35} + x_{65} + x_{75} - x_{53} - x_{54} - x_{56} \geq 170 \qquad \text{(Node 5)}$$
$$x_{56} + x_{76} - x_{65} \geq 70 \qquad \text{(Node 6)}$$
$$x_{76} + x_{75} + x_{74} \leq 300 \qquad \text{(Node 7)}$$

$$x_{ij} \geq 0 \quad \forall (i,j) \in E$$



Minimize:

$$\sum_{ij}\left(c_{ij}^{f} \cdot y_{ij}\right) + \sum_{ij}\left(c_{ij}^{v} \cdot x_{ij}\right) + \sum_{i}\left(c_{i}^{f} \cdot y_{i}\right) + \sum_{i}(p_i \cdot s_i) +$$

Where:

- $\sum_{ij}\left(c_{ij}^{f} \cdot y_{ij}\right)$: Edge fixed cost contribution
- $\sum_{ij}\left(c_{ij}^{v} \cdot x_{ij}\right)$: Variable cost contribution
- $\sum_{i}\left(c_{i}^{f} \cdot y_{i}\right)$: Node fixed cost contribution
- $\sum_{i}(p_i \cdot s_i)$: Penalty contribution
- $\sum_{ij}\left(l_{ij} \cdot x_{ij}\right)$: Edge lead time weighted by flow
- $\sum_{i} l_i \cdot \sum_{j} x_{ij}$: Node service time weighted by flow

s.t.

$$\sum_j x_{ji} - \sum_j x_{ij} = \text{ or } \leq \text{ or } \geq d_i \qquad \text{Flow Conservation}$$

$$b_{ij}^l \leq x_{ij} \leq b_{ij}^u \qquad \text{Lower \& Upper Flow Bound}$$

$$x_{ij} \leq M \cdot y_{ij} \qquad \text{Fixed Cost Route}$$

$$\sum_i (x_{ji} + x_{ij}) \leq M \cdot y_i \qquad \text{Fixed Cost Node}$$

$$\sum_i x_{ij} + s_j \geq d_j \qquad \text{Unmet Demand Penalty}$$

$$x_{ij} \geq 0 \quad \forall (i, j) \in E \qquad \text{Non Negative Flow}$$

Node Properties:

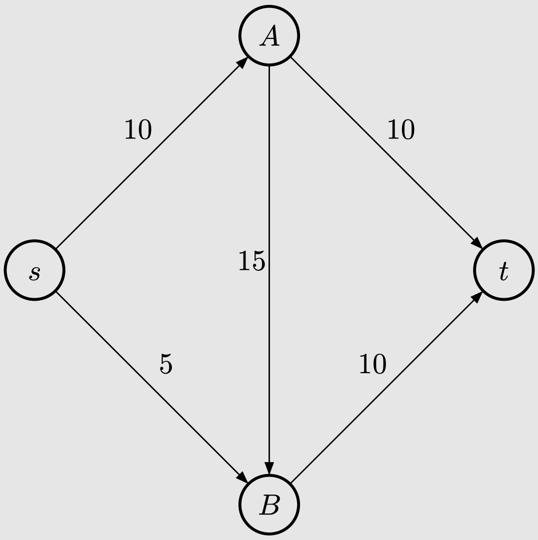
- **Node Type**: Source, sink, or intermediary.
- **Supply (Source)**: Flow capacity of source nodes.
- **Holding Cost (Intermediary)**: Inventory cost for stored goods.
- **Service Time**: Processing time at the node.
- **Demand**: Required flow at sink nodes.
- **Storage Capacity (Intermediary)**: Maximum amount of goods that can be held at a node
- **Penalty for Unfulfilled Demand (Sink)**: Cost for unmet demand.
- **Disruption Risk (All Nodes)**: Probability of a node being unavailable due to unforeseen circumstances

Edge Properties:

- **Edge Type**: Transport mode (air, water, road, rail).
- **Fixed Cost**: Cost incurred for using the edge, regardless of flow.
- **Reliability**: Probability of edge availability.
- **Flow Bounds**: Minimum and maximum allowable flow.
- **Unit Cost**: Cost per unit of flow.
- **Lead Time**: Time it takes for flow to travel along the edge.
- **Environmental Impact**: Account for the carbon footprint of using certain transport modes.

32.4. Ford-Fulkerson

Find augmenting paths in the network and increase the flow until no more augmenting paths can be found



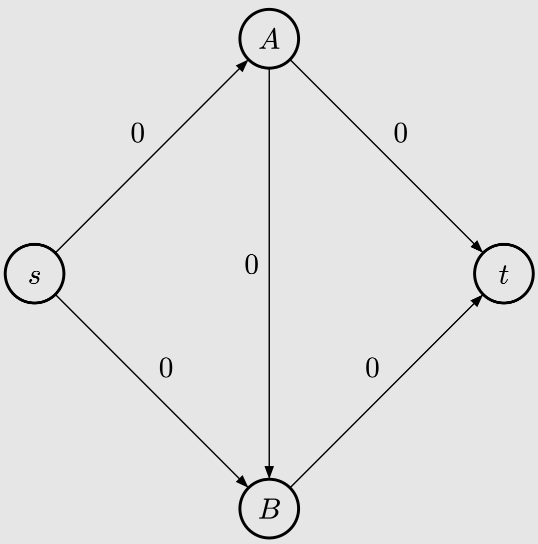
- s: Source
- A & B: Intermediate nodes
- t: Sink

Capacities:

- $s \to t$: 10
- $s \to B$: 5
- $A \to B$: 15
- $A \to t$: 10
- $B \to t$: 10

Step 1: Initialize flow to 0

All flows through the edges are initially set to 0.



Step 2: Find an augmenting path

Find an augmenting path using Depth-First Search (DFS). Start from the source s

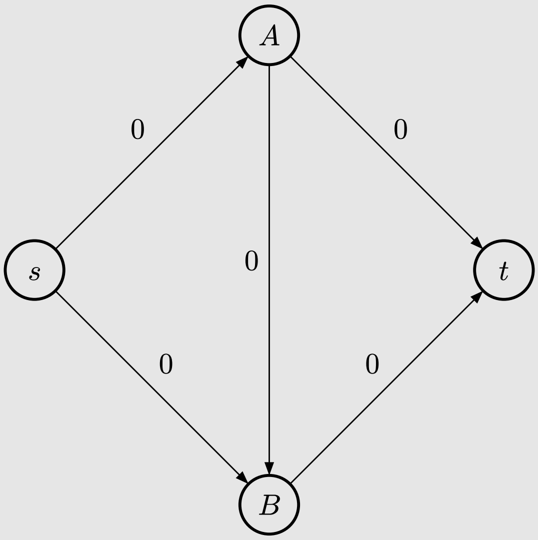
$$s \to A \to t$$

The minimum capacity along this path is 10 (bottleneck on edge $A \to t$)

We can push a flow of 10 units along this path.

Step 3: Update the residual graph

- $s \to A$: Capacity becomes $10 - 10 = 0$ (no residual capacity)
- $A \to t$: Capacity becomes $10 - 10 = 0$ (no residual capacity)



Step 4: Find another augmenting path

Find another augmenting path.

We cannot use $s \to A$ or $A \to t$ because their capacities are 0.

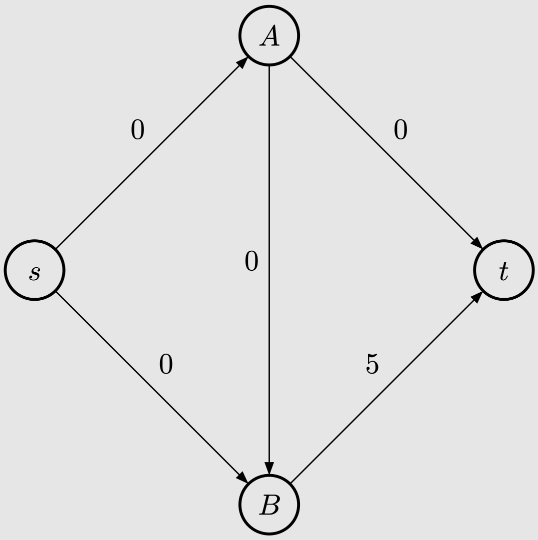
$$s \to B \to t$$

The minimum capacity along this path is 5 (bottleneck on edge $s \to B$)

We can push a flow of 5 units along this path

Step 5: Update the residual graph

- $s \to B$: Capacity becomes $5 - 5 = 0$
- $B \to t$: Capacity becomes $10 - 5 = 5$



Step 6: Termination

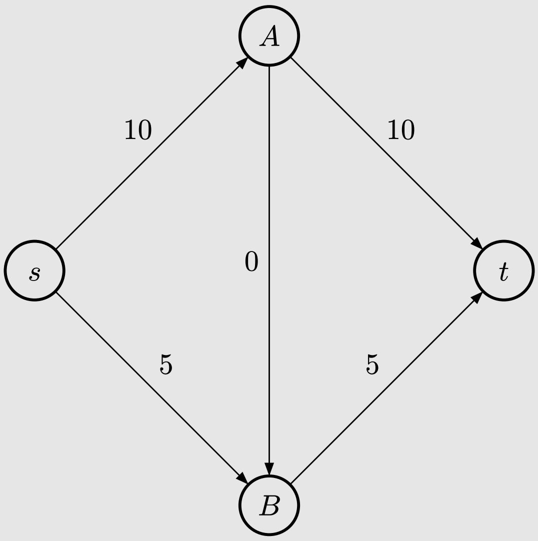
No more augmenting paths from s to t can be found, as all edges from s are fully saturated

Maximum flow:

- $s \to A \to t$: 10 units
- $s \to B \to t$: 5 units

Thus, the maximum flow from source s to sink t is 15 units

Step 7: Final Flow Distribution



33. Optimization

33.1. LP (Linear Programming)

Optimizing (maximizing or minimizing) a linear **objective function** subject to linear equality or inequality **constraints**. **Decision variables** can take any continuous real values.

1. Objective Function

Maximize:

$$Z = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

Or, equivalently:

$$Z = c^T x = \sum_{i=1}^{n} c_i x_i$$

Where:

- Z: Objective Function
- $c_1, c_2, \dots, c_n$: Coefficients
- $x_1, x_2, \dots, x_n$: Decision Variables

2. Constraints

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\le b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\le b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\le b_2 \end{aligned}$$

Or, in matrix form:

$$Ax \le b$$

Where:

- A: $m \times n$ matrix of coefficients a_{ij}
- $x = (x_1, x_2, \dots, x_n)^T$: vector of decision variables
- $b = (b_1, b_2, \dots, b_m)^T$: ecotr of known constants

3. Non-Nagativity Constraints

$$x_i \ge 0 \text{ for } i = 1, 2, \dots, n$$

1. Problem

A company produces two products: x_1 (Product A) and x_2 (Product B). The company wants to maximize profit, where:

- Each unit of Product A gives a profit of $40.
- Each unit of Product B gives a profit of $30.

The company has constraints on the production process:

- It takes 2 hours of labor to produce one unit of Product A and 1 hour to produce one unit of Product B. The company has a maximum of 100 labor hours available.
- The company can only use up to 80 units of raw material, and each unit of Product A uses 1 unit of material, while Product B uses 2 units of material.

The goal is to decide how many units of Product A (x_1) and Product B (x_2) to produce to maximize profit.

2. Formulation:

Objective Function (maximize the profit):

$$Z = 40x_1 + 30x_2$$

Constraints

1. Labor (maximum 100 hours):

$$2x_1 + x_2 \leq 100$$

2. Raw material (maximum 60 units):

$$x_1 + 2x_2 \leq 80$$

3. Non-negativity (can't produce nagtive quantities):

$$x_1 \geq 0$$

$$x_2 \geq 0$$

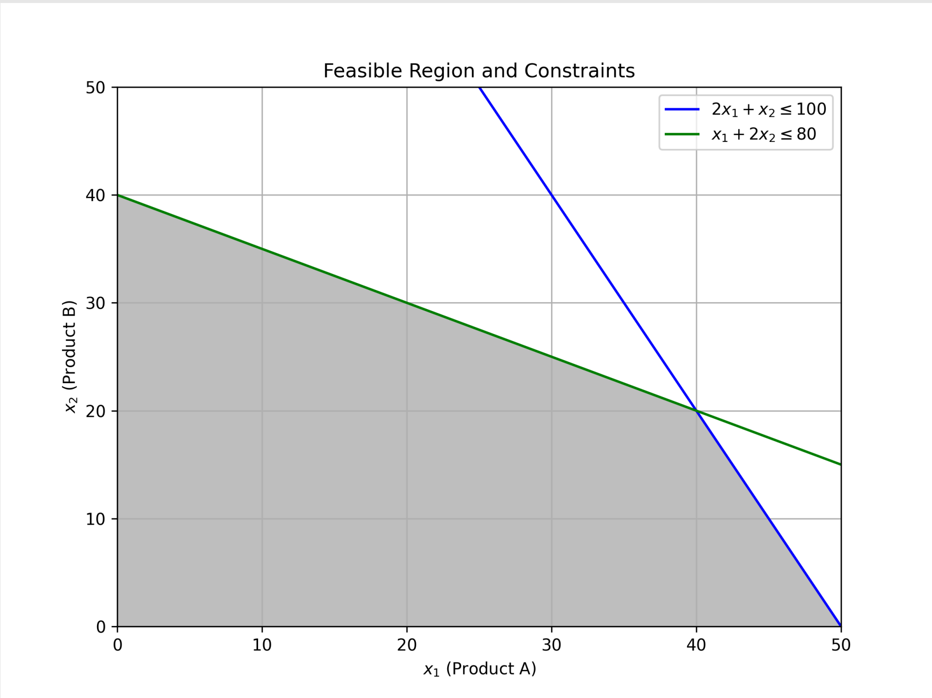
4. Summary

Maximize:

$$Z = 40x_1 + 30x_2$$

s.t.

$$\begin{aligned} 2x_1 + x_2 &\leq 100 \\ x_1 + 2x_2 &\leq 80 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned}$$




```python
import pulp

# Initialize the problem
prob = pulp.LpProblem("Maximize Profit", pulp.LpMaximize)
```

```python
# Define decision variables
x1 = pulp.LpVariable('x1', lowBound=0, cat='Continuous')  # Product A
x2 = pulp.LpVariable('x2', lowBound=0, cat='Continuous')  # Product B

# Objective function: Maximize 40*x1 + 30*x2
prob += 40 * x1 + 30 * x2, "Total Profit"

# Constraints
prob += 2 * x1 + x2 <= 100, "Labor Constraint"
prob += x1 + 2 * x2 <= 80, "Material Constraint"

# Solve the problem
prob.solve()

# Print the results
print("Status:", pulp.LpStatus[prob.status])
print(f"Optimal x1 (Product A): {pulp.value(x1)}")
print(f"Optimal x2 (Product B): {pulp.value(x2)}")
print(f"Maximum Profit: {pulp.value(prob.objective)}")
```

33.2. IP (Integer Programming)

Optimizing (maximizing or minimizing) a linear **objective function** subject to linear equality or inequality **constraints**. **Decision variables** can take any integer real values.

Minimize or Maximize

$$c^T x$$

s.t.

$$Ax \leq b$$

and

$$x \in \mathbb{Z}^n$$

Where

- x: vector of decision variables
- c: vector of coefficients for the objective function
- A: matrix of contraint coefficients
- b: vector of constraint constants
- $x \in \mathbb{Z}^n$: each x_i of x must be integer values

You are organizing a small event and want to minimize costs. You have to decide how many chairs and tables to rent.

- Each chair (x_1) costs \$5, each table (x_2) costs \$20.
- You need at least 3 tables and 10 chairs.
- Your budget is \$100.

You can only rent whole numbers of chairs and tables.

Minimize

$$5x_1 + 20x_2$$

s.t.

$$
\begin{aligned}
x_1 &\geq 10 \\
x_2 &\geq 3 \\
5x_1 + 20x_2 &\leq 100
\end{aligned}
$$

$$x_1, x_2 \in \mathbb{Z}^+$$

```python
import pulp

# Create a linear programming problem instance
# We are minimizing the cost
prob = pulp.LpProblem("Minimize_Cost", pulp.LpMinimize)

# Define decision variables
# x1 is the number of chairs
# x2 is the number of tables
x1 = pulp.LpVariable("x1", lowBound=10, cat='Integer')
x2 = pulp.LpVariable("x2", lowBound=3, cat='Integer')

# Objective function: Minimize 5*x1 + 20*x2
prob += 5 * x1 + 20 * x2, "Total_Cost"

# No additional constraints in this example, as the bounds cover the requirements

# Solve the problem
prob.solve()

# Print the results
print(f"Status: {pulp.LpStatus[prob.status]}")
print(f"Number of chairs (x1): {x1.varValue}")
print(f"Number of tables (x2): {x2.varValue}")
print(f"Total cost: {pulp.value(prob.objective)}")
```

33.3. Gradient Descent

Find the minimum of a function

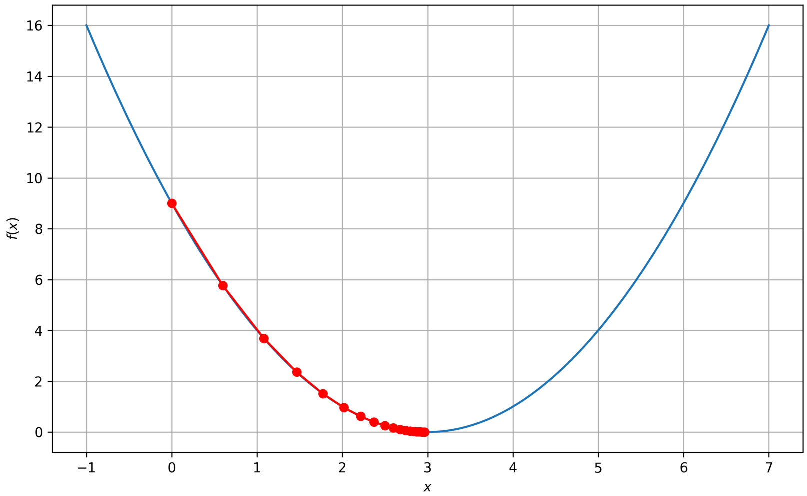
1. **Initialize**: Start with an initial guess for the parameters
2. **Compute Gradient**: Find the gradient of the function at the current parameters
3. **Update Parameters**: Adjust the parameters by moving in the opposite direction of the gradient, scaled by the learning rate
4. **Repeat**: Continue the process until the parameters converge to a minimum or the changes are minimal

Update Rule:

$$\theta \leftarrow \theta - \alpha \nabla f(\theta)$$

Where:

- θ: parameter being optimized
- α: learning rate
- $\nabla f(\theta)$: gradient of the funtion f with respect to θ



1. Function and Gradient:
    - Function: $\theta_1^2 + \theta_2^2$
    - Gradient: $\nabla f(\theta) = \left(\frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}\right) = (2\theta_1, 2\theta_2)$
2. Initial Values:

$$\theta_1 = 1$$
$$\theta_2 = 2$$

3. Learning Rate:

$$\alpha = 0.1$$

4. Gradient Calculation
    - For $\theta_1 = 1$ and $\theta_2 = 2$:

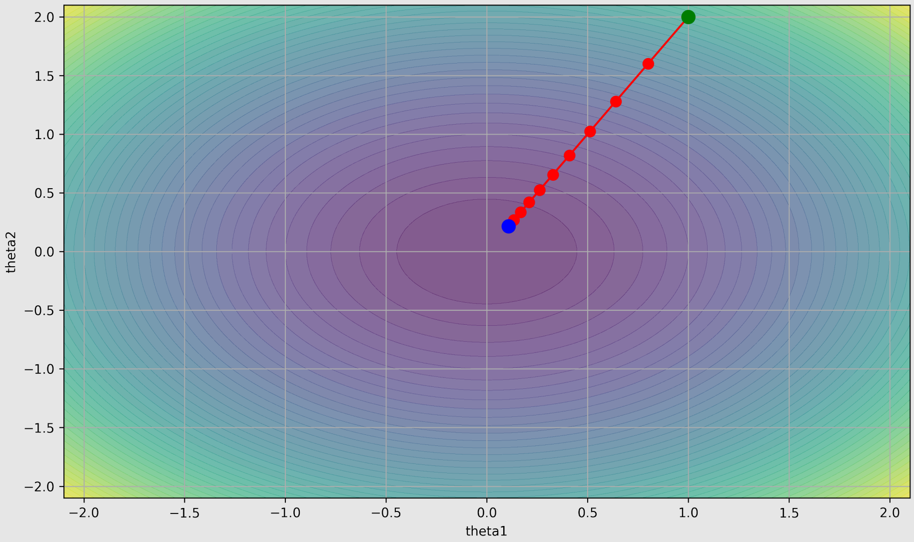
$$\nabla f(\theta) = (2 \cdot 1, 2 \cdot 2) = (2, 4)$$

5. Parameter Update:
    - Update θ_1 and θ_2 using the rule $\theta \leftarrow \theta - \alpha \nabla f(\theta)$:

$$\theta_1 \leftarrow 1 - 0.1 \cdot 2 = 1 - 0.2 = 0.8$$
$$\theta_2 \leftarrow 1 - 0.1 \cdot 4 = 2 - 0.4 = 1.6$$

6. New Values:

$$\theta_1 = 0.8$$
$$\theta_2 = 1.6$$



33.4. Monte Carlo