AIFS Lecture 2: Objects and Operations in Linear Algebra

Suraj Narayanan Sasikumar

Hessian AI Labs

March 01, 2020

Overview

Table of Contents

What is Linear Algebra?

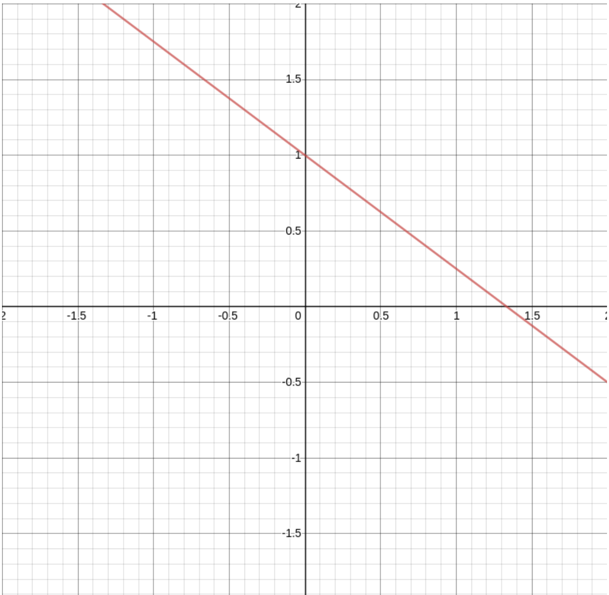
- Linear Algebra is the field of mathematics where linear equations and linear functions are represented by the interplay between scalars, vectors, matrices and vector-spaces.
- *Linear Equations*: An equation of the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b_1$$

  Here the variables, $x_1, x_2, \ldots$, does not have power greater than 1.
  Eg: $3x_1 + 4x_2 = 4,\ 3x_1 + 4x_2 + 5x_3 = 4$
- *Linear Functions*: A function $f : X \to Y$ is said to be a linear function if
  1. for all $n \in \mathbb{N}, f(x_1 + x_2 + \ldots x_n) = f(x_1) + f(x_2) + \ldots + f(X_n)$
  2. $af(x) = f(ax)$

2D-Linear equation

$3x_1 + 4x_2 = 4$



3D-Linear equation

$$3x_1 + 4x_2 + 5x_3 = 4$$

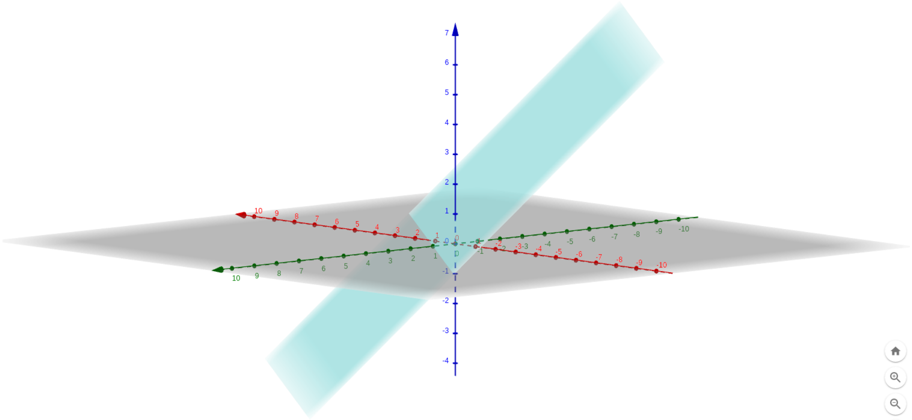


Table of Contents

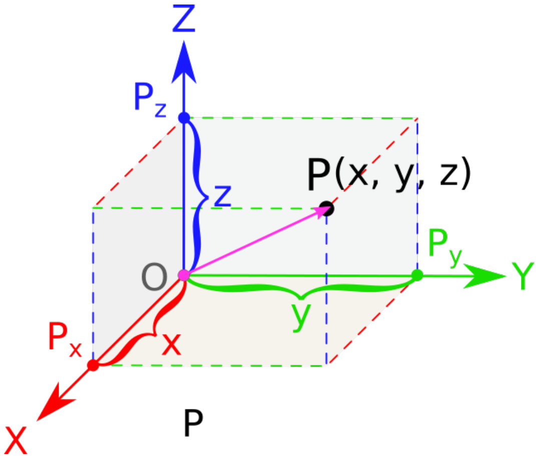
Scalars and Vectors

- *Scalar*: A single number. Usually written with lower case non-bold variable names. Eg: $3, 4.5, \frac{1}{3}, a$ for all $a \in \mathbb{N}$
- *Vector*: An array of ordered numbers. Can be written as a column of numbers, or as an n-tuple. Usually written with lower case bold variable names.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ OR } \mathbf{x} = (x_1, x_2, \ldots, x_n)$$

- A vector can be thought of as a 1-D array or numbers or a 2-D array with one column.

Cartesian Coordinate View

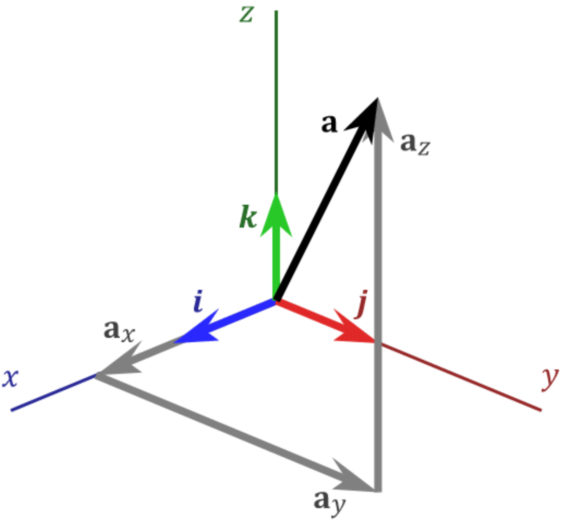


A vector can be visualized as an object identifying points in space, with each component of the vector giving the coordinate along a different axis.

Vector space

- A real valued vector space is a set V on which two operations $+$ and $\cdot$ are defined, called vector addition and scalar multiplication, and satisfy the following three properties for all scalars $c \in \mathbb{R}$.
- Closure under addition: If $\mathbf{u}$ and $\mathbf{v}$ are vectors in V then $\mathbf{u} + \mathbf{v}$ should also be in V.
- Closure under multiplication: If c is any scalar in $\mathbb{R}$ and $\mathbf{u}$ is a vector in V then $c \cdot \mathbf{u}$ is also in V.
- The zero vector $\mathbf{0}$ is in V.

Vector space in $\mathbb{R}^3$



Matrices

- A *matrix* is a 2-D array of numbers arranged in rows and columns for which the addition $(+)$ and multiplication $(\times)$ operations are defined.

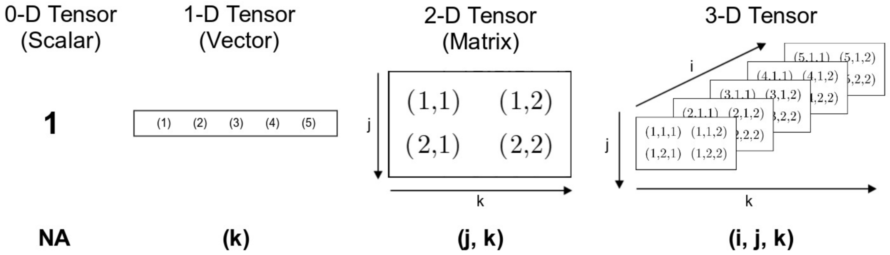
$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

- If a matrix $\boldsymbol{A}$ has m rows and n columns then we call it an $m \times n$ matrix, read as *"m by n matrix"* where m and n are called it's *dimensions*. $m \times n$ is also referred to as the *shape* of the matrix.
- *Notations*: $\boldsymbol{A}_{m\times n}$ specifies the dimensions of the matrix as subscript. $A_{i,j}$ refers to the value in the i^{th} row and j^{th} column.
- Eg. of a 2×3 matrix

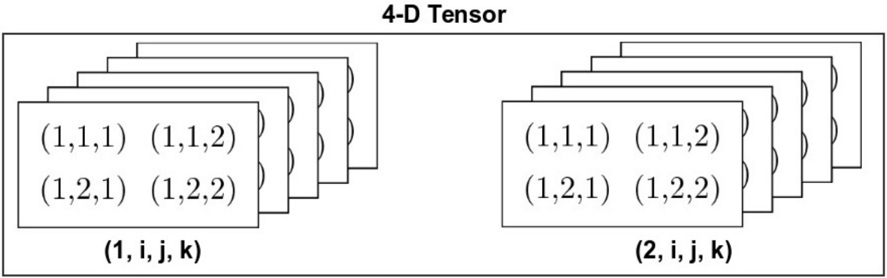
$$\boldsymbol{A}_{2\times 3} = \boldsymbol{A} = \begin{bmatrix} 5 & 9 & 1.2 \\ 3.5 & 5 & 6 \end{bmatrix} \text{ here, } A_{1,3} = 1.2$$

Tensors

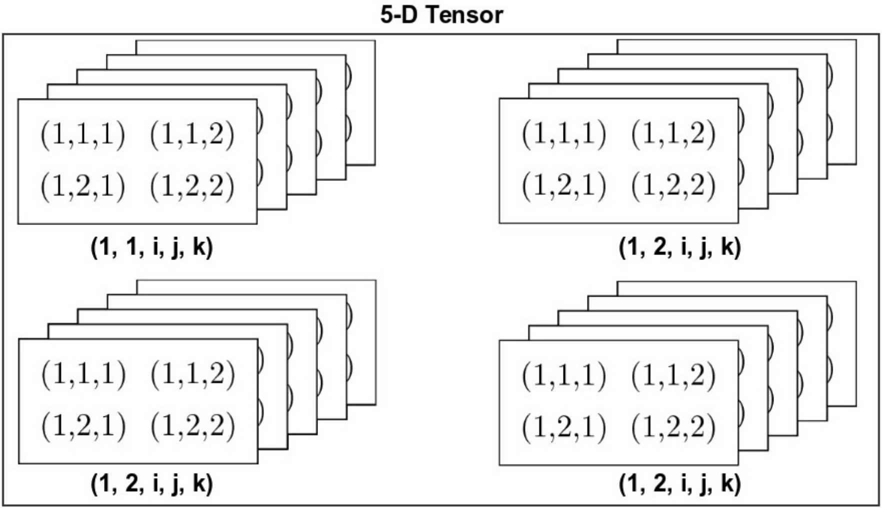
- A tensor is a mathematical object represented as a multidimensional array of numbers. It can be thought of as a generalization of the matrix to N-dimensions.



4D-Tensor



5D-Tensor

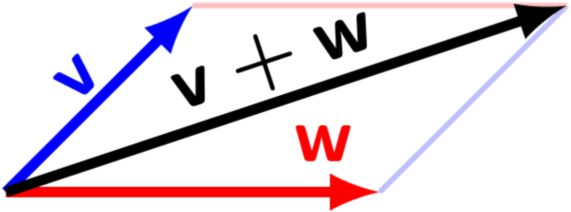


5-D Tensor
(1,1,1) (1,1,2)
(1,2,1) (1,2,2)
(1, 1, i, j, k)
(1,1,1) (1,1,2)
(1,2,1) (1,2,2)
(1, 2, i, j, k)
(1,1,1) (1,1,2)
(1,2,1) (1,2,2)
(1, 2, i, j, k)
(1,1,1) (1,1,2)
(1,2,1) (1,2,2)
(1, 2, i, j, k)

Table of Contents

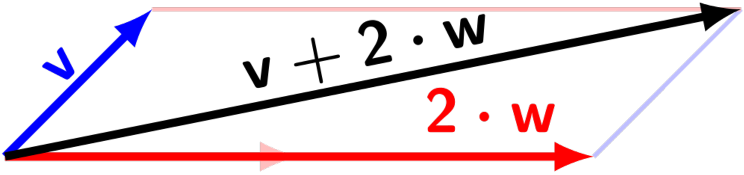
Vector addition

- Let $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$,
- then $\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n)$



Scalar Multiplication

- Let $\mathbf{w} = (w_1, w_2, \ldots, w_n)$, and c a scalar,
- then $c \cdot \mathbf{w} = (c \cdot w_1, c \cdot w_2, \ldots, c \cdot w_n)$



Matrix Addition

- *Matrix Addition*: Two matrices can only be added together if they have the same shape. Given two matrices $\boldsymbol{A}, \boldsymbol{B}$ of the same shape, then $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$, where $C_{i,j} = A_{i,j} + B_{i,j}$

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

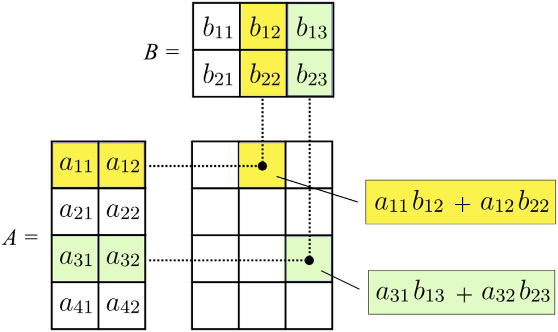
$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

- *Scalar Multiplication and Addition*: Given a matrix $\boldsymbol{A}$ and scalars c, d, the matrix $\boldsymbol{B} = c \cdot \boldsymbol{A} + d$ is given by $B_{i,j} = c \cdot A_{i,j} + d$

Matrix Multiplication

- *Matrix Multiplication*: Two matrices can only be multiplied together if the number of columns of the first matrix match the number of rows in the second. More concretely, given two matrices $\boldsymbol{A}_{m\times n}$ and $\boldsymbol{B}_{p\times q}$, the product $\boldsymbol{C} = \boldsymbol{AB}$ is only defined when $n = p$. The product is defined as:

$$C_{i,j} = \sum_{k=1}^{n} A_{i,k} B_{k,j}$$



Transpose Operation

- Transpose of a matrix (or a vector represented as a matrix) is the mirror image of the matrix across its main diagonal.
- Easier way to visualize: rows become columns and columns become rows.
- Transpose of a matrix $\boldsymbol{A}$ is written as $\boldsymbol{A}^\top$

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

Trace Operator

Trace is a matrix operator that returns the sum of the elements along the main diagonal (the diagonal starting from top-left moving right and down one row and column until we reach the bottom row) of a square matrix.

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

$$\mathrm{Tr}(\boldsymbol{A}) = \sum_i A_{i,i} = a_{11} + a_{22} + \ldots + a_{nn}$$

Determinant

- The determinant of a matrix is a scalar value computed from the elements of a square matrix.
- It is denoted by $\det(\boldsymbol{A})$ or $|\boldsymbol{A}|$, and given by the formula (not important).

$$\det(\boldsymbol{A}) = \sum_{\sigma \in S_n} \left(sgn(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i} \right)$$

In the case of a 2 × 2 matrix the determinant may be defined as

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Similarly, for a 3 × 3 matrix A, its determinant is

$$\begin{aligned} |A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a \begin{vmatrix} \square & \square & \square \\ \square & e & f \\ \square & h & i \end{vmatrix} - b \begin{vmatrix} \square & \square & \square \\ d & \square & f \\ g & \square & i \end{vmatrix} + c \begin{vmatrix} \square & \square & \square \\ d & e & \square \\ g & h & \square \end{vmatrix} \\ &= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= aei + bfg + cdh - ceg - bdi - afh. \end{aligned}$$

Table of Contents

Row and Column Vectors

A vector can be written in matrix format in two ways: as row vectors and column vectors.

- *Column-vector*: A single column with more than one row, that is a matrix with m rows and 1 column.
- *Row-vector*: A single row with more than one columns, that is a matrix with 1 row and n columns.

$$\boldsymbol{r} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \end{bmatrix} \qquad \boldsymbol{c} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

Square Matrix

An $m \times n$ matrix where $m = n$, that is the number of rows is equal to the number of columns.

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

$$\text{Eg: } \boldsymbol{A} = \begin{bmatrix} 3 & 4 & 9 & 1 \\ 7 & 4 & 1 & 8 \\ 7 & 3 & 2 & 9 \\ 5 & 6 & 9 & 0 \end{bmatrix}$$

Symmetric Matrix

A Symmetric matrix is a square matrix where $A = A^T$

$$\text{Eg: } \boldsymbol{A} = \begin{bmatrix} 3 & 4 & 9 & 1 \\ 4 & 4 & 3 & 6 \\ 9 & 3 & 2 & 1 \\ 1 & 6 & 1 & 0 \end{bmatrix}$$

Diagonal Matrix

- A **diagonal matrix** is a matrix where all elements outside the main diagonal (the diagonal starting from top-left moving right and down one row and column until we reach the bottom row).
- A diagonal matrix can also be a rectangular matrix, but in most cases it would be a **square symmetric matrix**.

$$\text{Eg: } \boldsymbol{A} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix}$$

Identity Matrix

- An **identity matrix** is a square matrix with ones on the main diagonal and zeros everywhere else.
- The letter $\boldsymbol{I}_n$ is used to denote an $n \times n$ identity matrix.

$$\text{Eg: } \boldsymbol{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Invertible Matrix

- An $n \times n$ square matrix, $\boldsymbol{A}$, is said be **invertible** if there exists another $n \times n$ square matrix, $\boldsymbol{B}$, such that,

$$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}_n$$

- If the matrix $\boldsymbol{B}$ exists, it is written as $\boldsymbol{A}^{-1}$
- The inverse of a matrix, $\boldsymbol{A}$ exists if and only if $\det(\boldsymbol{A}) \neq 0$
- A non-invertible matrix is also known as a *singular* matrix

$$\text{Eg: } \boldsymbol{A} = \begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix} \quad \boldsymbol{A}^{-1} = \begin{bmatrix} -1 & \frac{3}{2} \\ 1 & -1 \end{bmatrix}$$

$$\boldsymbol{AA}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{I}_2$$

Table of Contents

More about Spaces

- With vector spaces we added the notions of addition and scalar-multiplication of vectors. But what about...
- Distance between two points (vectors)
- Length (magnitude, size) of a vector
- Angle between two vectors

More about Spaces

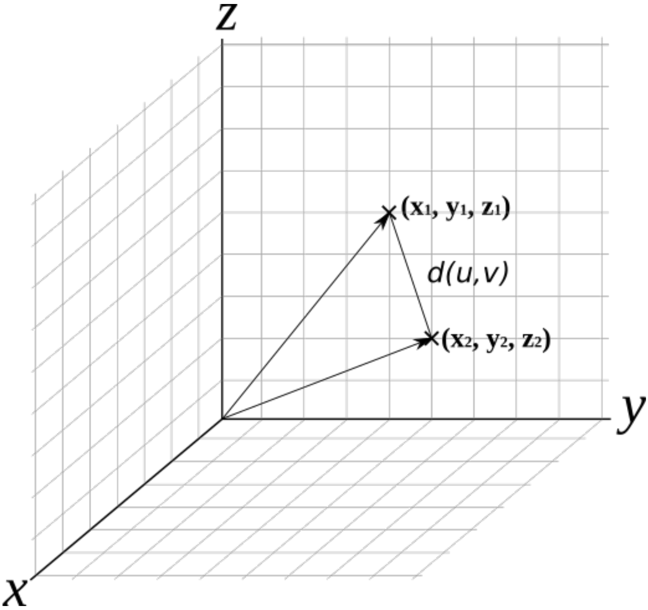
- With vector spaces we added the notions of addition and scalar-multiplication of vectors. But what about...
- Distance between two points (vectors)... Metric space
- Length (magnitude, size) of a vector... Normed Space
- Angle between two vectors... Inner Product Space

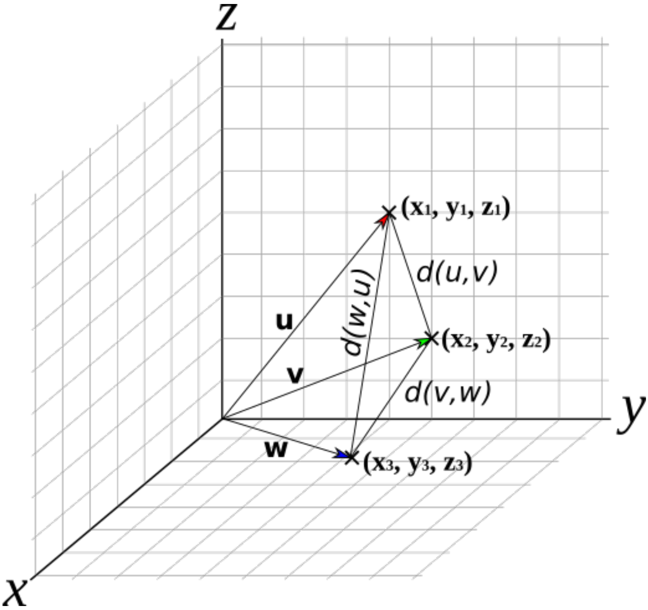
Metric Spaces

- A vector space with a *distance function* giving the distance between any two points is called a metric space. [1]
- For a function, $d(\cdot, \cdot)$, to be a distance function, it is needs to satisfy the following properties.
- *Identity of Indiscernibles*: $d(\mathbf{u}, \mathbf{v}) = 0 \Rightarrow \mathbf{u} = \mathbf{v}$
- *Symmetry*: $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$
- *Triangle Inequality*: $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$

[1] This definition is correct but not complete, to flush out the actual details we need to go more into pure math, which is not required for us.

Metric Spaces: Visualization




z
y
x
(x1, y1, z1)
(x2, y2, z2)
(x3, y3, z3)
d(u,v)
d(w,u)
d(v,w)
u
v
w

Examples of Metric Spaces

- $\mathbb{R}^n$ with euclidean distance. If $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$,

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \ldots + (u_n - v_n)^2}$$

- Manhattan distance or taxi-cab distance or L_1-distance. If $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$,

$$d(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} |u_i - v_i|$$

- Wasserstein metric. A metric that gives a measure of distance between two probability distributions. Useful in Deep learning. Formula!

More about Spaces

- With vector spaces we added the notions of addition and scalar-multiplication of vectors. But what about...
- Distance between two points (vectors)... Metric space $(V, +, \cdot, d(\cdot, \cdot))$
- Length (magnitude, size) of a vector... Normed Space
- Angle between two vectors... Inner Product Space

Normed Space

- A *Norm* is a function that gives the length (magnitude/size) of a vector. The notation $||\cdot||$ is often used for a norm.
- Since a vector is representative of a point in space, the length of a vector is always relative to the origin (the zero vector: $\mathbf{0}$)
- For a function, $f(\cdot) = ||\cdot||$, to be a norm, it has to satisfy the following properties.
- *Triangle Inequality*: $||\mathbf{u}+\mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$
- *Absolute Homogeneity*: $||a\mathbf{u}|| = a||\mathbf{u}||$
- *Positive Definite*: If $||\mathbf{u}|| = 0 \Rightarrow \mathbf{u} = 0$

Examples of Normed spaces

- $\mathbb{R}^n$ with ℓ_1-*norm* (Taxicab norm):

$$||\mathbf{u}|| = ||\mathbf{u}||_1 = \sum_{i=1}^{n} |u_i|$$

- $\mathbb{R}^n$ with ℓ^2-*norm* (Euclidean norm):

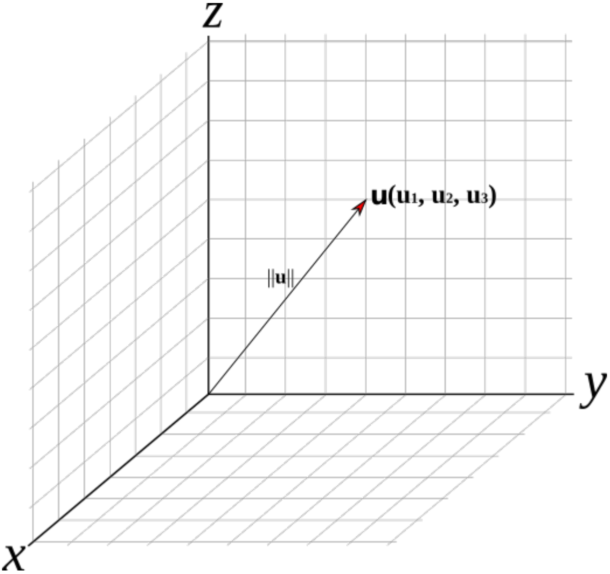
$$||\mathbf{u}|| = ||\mathbf{u}||_2 = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}$$

- $\mathbb{R}^n$ with *max-norm* (Infinity norm):

$$||\mathbf{u}|| = ||\mathbf{u}||_\infty = \max(|u_1|, |u_2|, \ldots, |u_n|)$$

Visualization: Normed Spaces

Do you a see a connection between metric spaces and normed spaces?

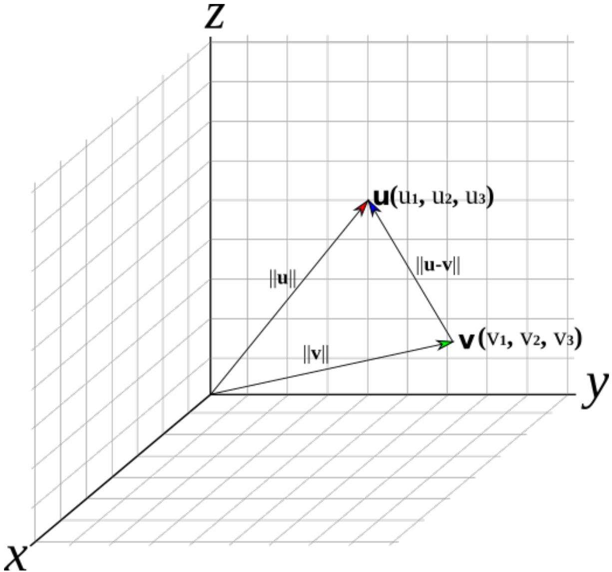


Norm induces a Metric

- A Norm induces a Metric in the space.
- If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$ and $||\mathbf{u}||$, $||\mathbf{v}||$ are their respective norms, then the norm $||\cdot||$ induces the following metric:

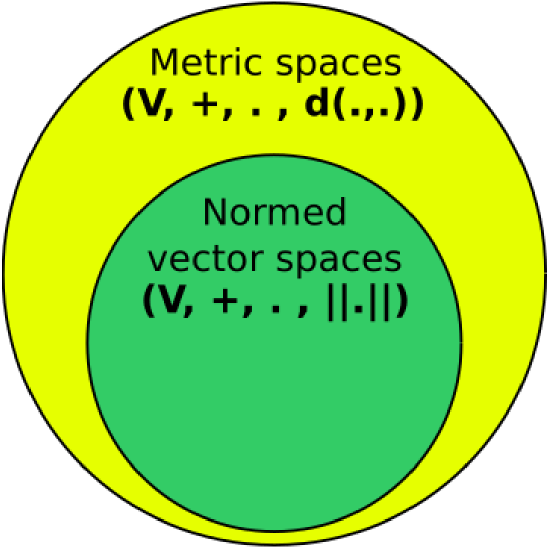
$$d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$$

Visualization: Norms and Metrics



Visualization: Norms and Metrics

All normed spaces are metric spaces, but not all metric spaces are normed spaces.



More about Spaces

- With vector spaces we added the notions of addition and scalar-multiplication of vectors. But what about...
- Distance between two points (vectors)... Metric space $(V, +, \cdot, d(\cdot, \cdot))$
- Length (magnitude, size) of a vector... Normed Space $(V, +, \cdot, || \cdot ||)$
- Angle between two vectors... Inner Product Space

Inner product space

- A vector space with an additional function called the *inner product*, is called an inner product space.
- An inner product, denoted by $\langle \cdot, \cdot \rangle$, is a function that takes two vectors and returns a scalar, and also satisfies the following properties.
- *Symmetry*: $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
- *Linearity in first (or second) argument*:
  $a\langle \mathbf{x}, \mathbf{y} \rangle = \langle a\mathbf{x}, \mathbf{y} \rangle$
  $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
- *Positive Definiteness*: for all $\mathbf{x} \in V \setminus \{\mathbf{0}\}$ $\langle \mathbf{x}, \mathbf{x} \rangle > 0$

Inner product space

- A vector space with an additional function called the *inner product*, is called an inner product space.
- An inner product, denoted by $\langle \cdot, \cdot \rangle$, is a function that takes two vectors and returns a scalar, and also satisfies the following properties.
  1. *Symmetry*: $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
  2. *Linearity in first (or second) argument*:
     $a\langle \mathbf{x}, \mathbf{y} \rangle = \langle a\mathbf{x}, \mathbf{y} \rangle$
     $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
  3. *Positive Definiteness*: for all $\mathbf{x} \in V \setminus \{\mathbf{0}\}$ $\langle \mathbf{x}, \mathbf{x} \rangle > 0$
- Important Results:
  - $\langle \mathbf{0}, \mathbf{0} \rangle = 0$

$$\langle \mathbf{0}, \mathbf{0} \rangle = \langle 0\mathbf{x}, 0\mathbf{x} \rangle = 0\langle \mathbf{x}, 0\mathbf{x} \rangle = 0 \qquad \text{(Linearity in first argument)}$$

  - *Cauchy–Schwarz inequality*:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}\sqrt{\langle \mathbf{y}, \mathbf{y} \rangle}$$

Inner Product induces a Norm

Theorem

$\sqrt{\langle x, x \rangle}$ *is a norm.*

Proof.

We have to prove that $\sqrt{\langle x, x \rangle}$ satisfies all the properties of a norm.

- *Positive Definiteness*: We need to prove $\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = 0 \to \mathbf{x} = \mathbf{0}$
  To do that let's prove the contrapositive, meaning it is enough to show that $\mathbf{x} \neq \mathbf{0} \to \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \neq 0$

$$\begin{aligned} \mathbf{x} \neq \mathbf{0} &\to \langle \mathbf{x}, \mathbf{x} \rangle > 0 && \text{(Positive Definiteness of Inner product)} \\ &\to \langle \mathbf{x}, \mathbf{x} \rangle \neq 0 \\ &\to \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \neq 0 \end{aligned}$$

$\square$

Inner Product induces a Norm

Proof.

- *Absolute Homogeneity*: We need to prove $\sqrt{\langle a\mathbf{x}, a\mathbf{x}\rangle} = |a|\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$

$$
\begin{aligned}
\sqrt{\langle a\mathbf{x}, a\mathbf{x}\rangle} &= \sqrt{a\langle \mathbf{x}, a\mathbf{x}\rangle} && \text{(Linearity in first argument)} \\
&= \sqrt{a^2\langle \mathbf{x}, \mathbf{x}\rangle} && \text{(Linearity in second argument)} \\
&= |a|\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle} && (\sqrt{a^2} = |a|)
\end{aligned}
$$

$\square$

Proof.

- *Triangle Inequality*: To show: $\sqrt{\langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle} \leq \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle} + \sqrt{\langle \mathbf{y}, \mathbf{y}\rangle}$

$$\begin{aligned}
\langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle &= \langle \mathbf{x}, \mathbf{x}+\mathbf{y}\rangle + \langle \mathbf{y}, \mathbf{x}+\mathbf{y}\rangle && \text{(Linearity in } 1^{st}\text{)} \\
&= \langle \mathbf{x}, \mathbf{x}\rangle + \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{y}, \mathbf{x}\rangle + \langle \mathbf{y}, \mathbf{y}\rangle && \text{('' in } 2^{nd}\text{)} \\
&= \left(\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}\right)^2 + 2\langle \mathbf{x}, \mathbf{y}\rangle + \left(\sqrt{\langle \mathbf{y}, \mathbf{y}\rangle}\right)^2 && \text{(Symmetry)} \\
&\leq \left(\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}\right)^2 + 2|\langle \mathbf{x}, \mathbf{y}\rangle| + \left(\sqrt{\langle \mathbf{y}, \mathbf{y}\rangle}\right)^2 && (a \leq |a|) \\
&\leq \left(\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}\right)^2 + 2\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}\sqrt{\langle \mathbf{y}, \mathbf{y}\rangle} + \left(\sqrt{\langle \mathbf{y}, \mathbf{y}\rangle}\right)^2 \\
& && \text{(Cauchy–Schwarz inequality)} \\
&= \left(\sqrt{\langle \mathbf{x}, \mathbf{x}\rangle} + \sqrt{\langle \mathbf{y}, \mathbf{y}\rangle}\right)^2 \\
\therefore \sqrt{\langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle} &\leq \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle} + \sqrt{\langle \mathbf{y}, \mathbf{y}\rangle}
\end{aligned}$$

$\square$

Inner Product induces a Norm and Metric

Definition

Give a vector **u** in an inner product space: $(V, +, \cdot, \langle \cdot, \cdot \rangle)$, the norm of the vector, $||\mathbf{u}||$ is defined as:

$$||\mathbf{u}|| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$$

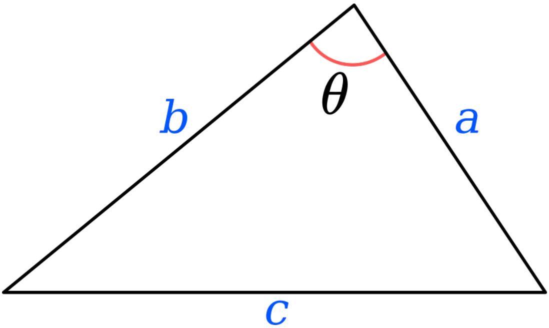
We know that a norm induces a metric in the space, therefore by association an inner product also induces a metric.

Definition

Give two vectors **u**, **v** in an inner product space: $(V, +, \cdot, \langle \cdot, \cdot \rangle)$, the distance between the vector, $d(\mathbf{u}, \mathbf{v})$, is defined as:

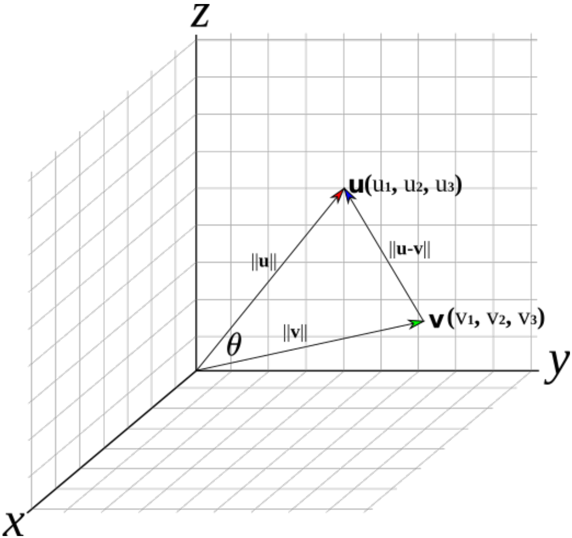
$$d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}|| = \sqrt{\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle}$$

Law of Cosines in Trigonometry



Law of Cosines: $c^2 = a^2 + b^2 - 2ab\cos\theta$

Angle between two Vectors



$$||\mathbf{u} - \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2 - 2||\mathbf{u}||.||\mathbf{v}|| \cos\theta$$

Angle between two Vectors

$$
\begin{aligned}
||\mathbf{u}-\mathbf{v}||^2 &= \left(\sqrt{\langle \mathbf{u}-\mathbf{v}, \mathbf{u}-\mathbf{v}\rangle}\right)^2 \\
&= \langle \mathbf{u}-\mathbf{v}, \mathbf{u}-\mathbf{v}\rangle \\
&= \langle \mathbf{u}, \mathbf{u}-\mathbf{v}\rangle + \langle -\mathbf{v}, \mathbf{u}-\mathbf{v}\rangle && \text{(Linearity in } 1^{st}\text{)} \\
&= \langle \mathbf{u}, \mathbf{u}\rangle - \langle \mathbf{u}, \mathbf{v}\rangle - \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle && \text{('' in } 2^{nd}\text{)} \\
&= \langle \mathbf{u}, \mathbf{u}\rangle - 2\langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle && \text{(Symmetry)}
\end{aligned}
$$

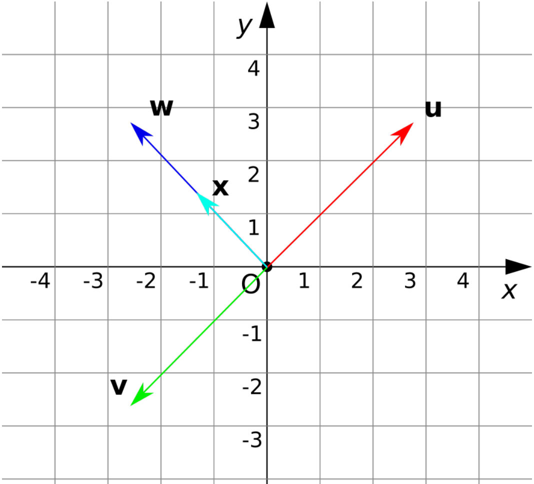
$$
\begin{aligned}
\therefore \langle \mathbf{u}, \mathbf{u}\rangle - 2\langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle &= ||\mathbf{u}||^2 + ||\mathbf{v}||^2 - 2||\mathbf{u}||.||\mathbf{v}|| \cos\theta \\
\Rightarrow \langle \mathbf{u}, \mathbf{u}\rangle - 2\langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle - 2||\mathbf{u}||.||\mathbf{v}|| \cos\theta \\
\Rightarrow \cos\theta &= \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{||\mathbf{u}||.||\mathbf{v}||} \\
\therefore \theta &= \arccos \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{||\mathbf{u}||.||\mathbf{v}||}
\end{aligned}
$$

Question?

Why did we need the inner product space if the angle can be derived from the Norm?

Angle between two Vectors

The previous derivation overlooked a special case. What happens when one vector is a scalar multiple of the other? The Law of cosines does not work anymore because the distance between the two vectors is not the side opposite to the angle. We need a new proof for such cases.



Angle between two Vectors

Give two vectors **u** and **v**, which are scalar multiples of each other, without loss of generality, we can write:

$$\mathbf{u} = c\mathbf{v}$$

The inner product between the two vectors can be written as:

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{v} \rangle &= \langle c\mathbf{v}, \mathbf{v} \rangle \\
&= c\langle \mathbf{v}, \mathbf{v} \rangle && \text{(Linearity in } 1^{st}\text{)} \\
&= \text{sign}(c)|c|\langle \mathbf{v}, \mathbf{v} \rangle \\
&= \text{sign}(c)|c|.||\mathbf{v}||^2 && (||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}) \\
&= \text{sign}(c)|c|.||\mathbf{v}||.||\mathbf{v}|| \\
&= \text{sign}(c)||c\mathbf{v}||.||\mathbf{v}|| && \text{(Absolute Homogeneity)} \\
\therefore \langle \mathbf{u}, \mathbf{v} \rangle &= \text{sign}(c)||\mathbf{u}||.||\mathbf{v}|| && (\mathbf{u} = c\mathbf{v})
\end{aligned}$$

Angle between two vectors

We have:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \text{sign}(c)||\mathbf{u}||.||\mathbf{v}||$$

When $\text{sign}(c) = 1, \theta = 0 \Rightarrow \cos\theta = \cos 0 = 1$ (Same Direction)
When $\text{sign}(c) = -1, \theta = 180 \Rightarrow \cos\theta = \cos 180 = -1$ (Opp. Direction)
Therefore, we can replace $\text{sign}(c)$ with $\cos\theta$.

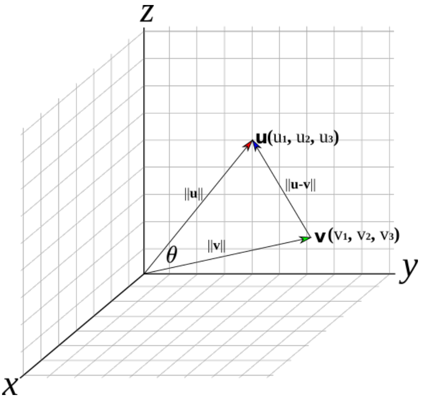
$$\Rightarrow \langle \mathbf{u}, \mathbf{v} \rangle = \cos\theta||\mathbf{u}||.||\mathbf{v}||$$

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}||.||\mathbf{v}||}$$

$$\therefore \theta = \arccos\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}||.||\mathbf{v}||}$$

Putting it all together

Given an **Inner Product Space**: $(V, +, ., \langle \cdot, \cdot \rangle)$. We have,



1. Angle between two vectors: $\theta = \arccos \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}||.||\mathbf{v}||}$
2. Length of a vector: $||\mathbf{u}|| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$
3. Distance between two vectors: $d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}|| = \sqrt{\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle}$

Dot Product

The dot product is the most famous example of an inner product. A Vector space equipped with the dot product is an inner product space. The dot product is defined as follows

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

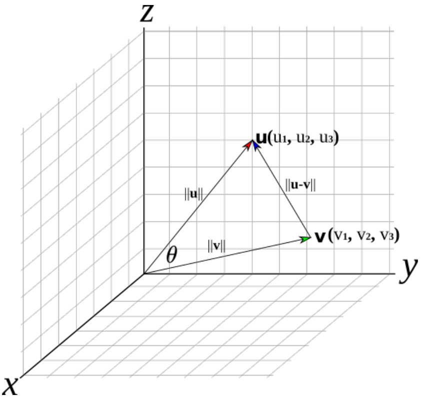
Where,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \text{ and } \mathbf{x}^T \mathbf{y} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\therefore \mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = [x_1 y_1 + x_2 y_2 + \dots x_n y_n]$$

Vector Space as we know it

Given a vector space, $(V, +, .)$ along with the dot product $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v}$.



1. Angle between two vectors: $\theta = \arccos \frac{\mathbf{u}^T\mathbf{v}}{||\mathbf{u}||.||\mathbf{v}||}$
2. Length of a vector: $||\mathbf{u}|| = \sqrt{\mathbf{u}^T\mathbf{u}}$
3. Distance b/w two vectors: $d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}|| = \sqrt{(\mathbf{u} - \mathbf{v})^T(\mathbf{u} - \mathbf{v})}$

Inner Product Spaces

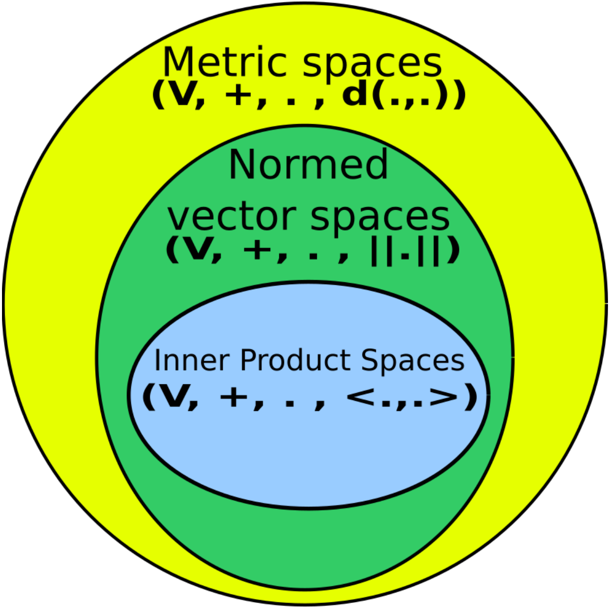


Table of Contents

Operations in Linear Algebra

From that not so brief aside we now have three new operations.

- *Dot Product*: An operation between two vectors that results in a scalar value. It is also know as the scalar product.

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \dots x_ny_n \qquad \text{(Component-wise)}$$

$$\mathbf{x} \cdot \mathbf{y} = ||\mathbf{x}||\,||\mathbf{y}||\cos(\theta) \qquad \text{(Using Angle)}$$

- *Norm*: The norm gives the magnitude or length of a vector. $||\mathbf{x}|| = \sqrt{\mathbf{x} \cdot \mathbf{y}} = \sqrt{\mathbf{x}^T\mathbf{y}}$
- *Distance*: Measures the distance between two vectors. $d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||$