VISPAD INSTITUTE OF TECHNOLOGY

Nea onnim no sua a ohu,

# Inner Product, Orthogonality, Norm

Fall 2025

(October – December Virtual Internship)


*Nea onnim no sua a ohu,*

# Program Timeline

10th October – 14th December

(10 Weeks)

# Lectures

**Sundays : 7:00 pm – 9:00 pm**

# Labs


**Saturdays : 7:00 pm – 9:00 pm**

**At the end of the session the students or candidates should be able to understand and work with:**

- Inner Products
- Inequalities
  - The Cauchy-Schwarz Inequality
  - The triangle Inequality
- Orthogonal Vectors and Orthogonal Bases
- Orthogonal Projection and the Closes Point
- The Gram-Schmidt Process
- Orthogonal Subspaces and Complements
- Norms

# RECAP

# Magnitude and Direction

- The **magnitude** of a vector, often called its **norm**, quantifies its length or size.
- In machine learning, the magnitude of a vector can sometimes indicate the intensity or importance of the represented data point or feature set.

- The **direction** of a vector tells us “where it points” in the vector space.
- Direction is independent of magnitude.
- The vectors $[1, 1]$ and $[3,3]$ point in the same direction, but different magnitude.
- Direction is fundamental for understanding relationships between vectors such as similarity.

- A ***unit vector*** is simply a vector with a magnitude of 1.
- Any non-zero vector $\boldsymbol{v}$ can be converted into a unit vector $u$ that points in the same direction by dividing the vector by its magnitude.
- $u = \frac{v}{\|v\|}$, this process is called normalization

- Let's normalize the vector $v = [3,4]$

$$\|v\| = \sqrt{x^2 + y^2} = \sqrt{3^2 + 4^2} = 5$$

$$u = \frac{v}{\|v\|} = \left[\frac{3}{5}, \frac{4}{5}\right] = [0.6, 0.8]$$

- The magnitude of a unit vector is 1.

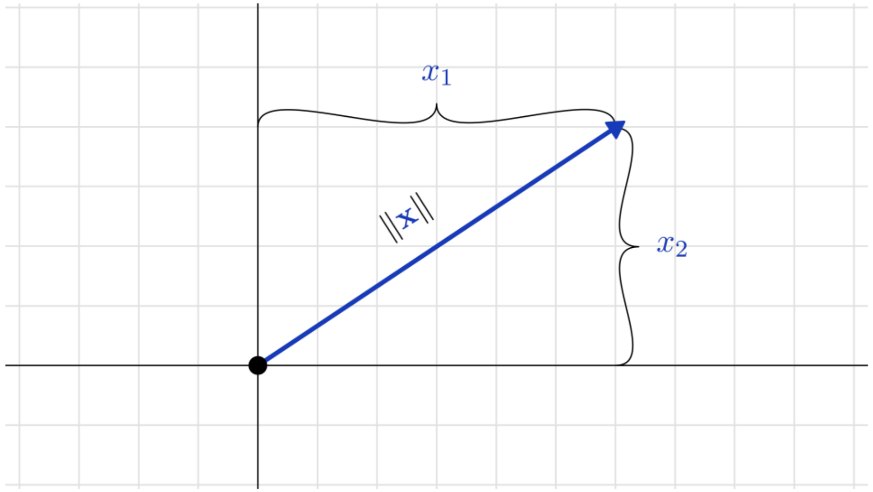
# Vector Norms: Measuring Length

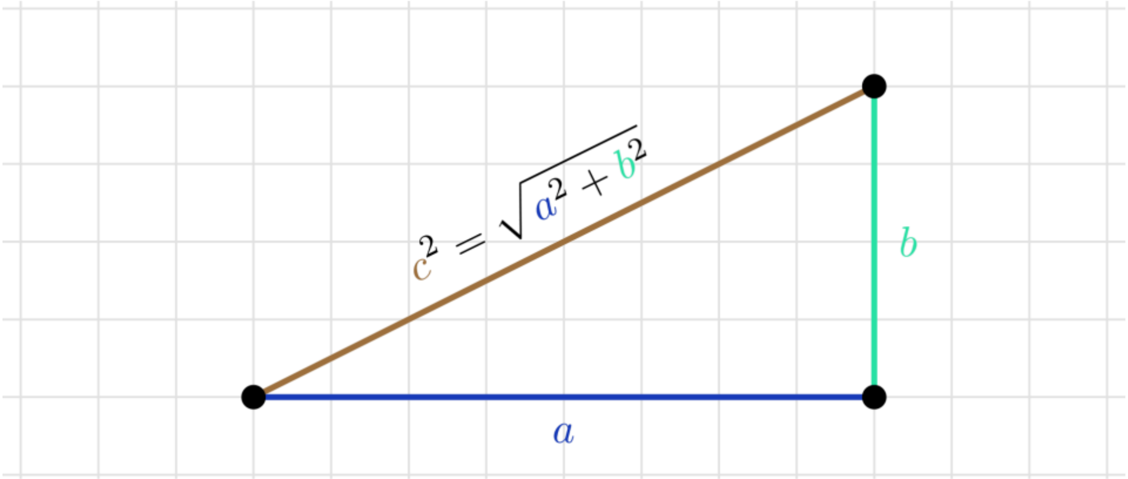
- The most familiar norm is the $L_2$ **norm**, also known as the **Euclidean norm.**
- It is defined as:

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2}, \qquad x = (x_1, x_2)$$

❑The Euclidean norm formula can be generalized to higher dimension by:

$$\|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}, \qquad \boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}$$



- The $\boldsymbol{L_1}$ **norm** also called the **Manhattan norm** or Taxicab norm sums the absolute values of the individual component of the vector.

$$\|x\|_1 = |x_1| + |x_2|, \qquad x = (x_1, x_2)$$

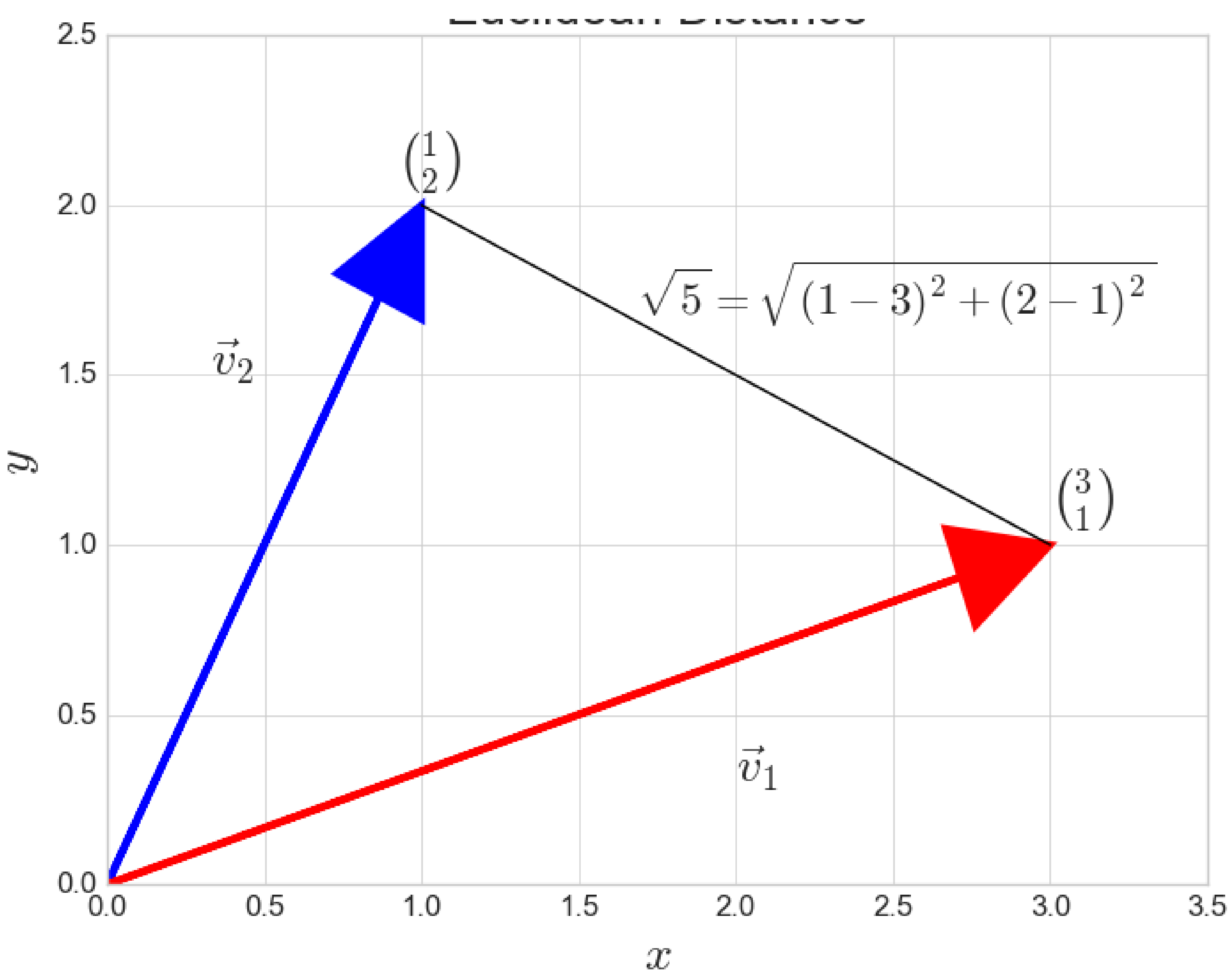
The $L_1$ norm can be also be generalized to a higher dimension by:

$$\|x\|_1 = |x_1| + \cdots + |x_2|, \qquad \boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}$$

- The $L_2$ **norm** squares values, heavily penalizing large components.
- The $L_1$ **norm** takes absolute values, treating deviations linearly.

# Vector Distance

- The most common way to measure the distance between two vectors is the **Euclidean distance.**
- It represents the straight-line distance between two points defined by the vectors in the feature space.



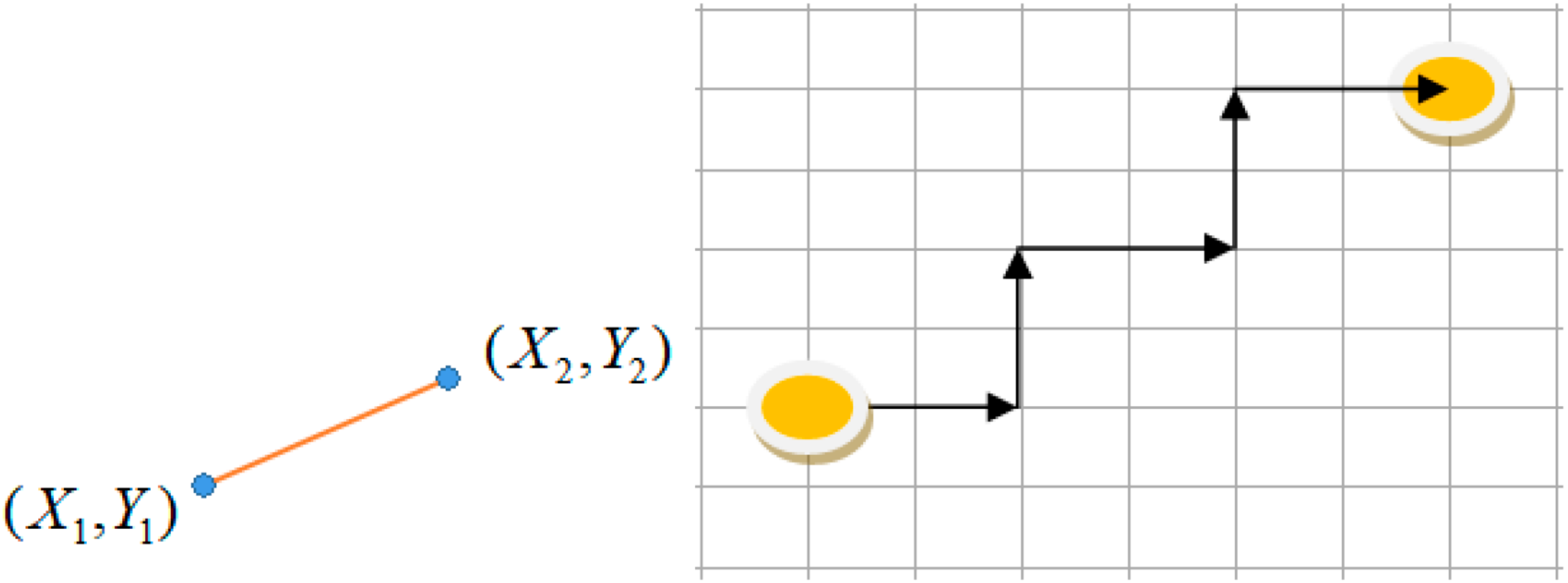
$\binom{1}{2}$
$\binom{3}{1}$
$\sqrt{5} = \sqrt{(1-3)^2 + (2-1)^2}$
$\vec{v}_2$
$\vec{v}_1$
$x$
$y$
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5

$\boldsymbol{u} = (u_1, u_2, .., u_n)$ and $\boldsymbol{v} = (v_1, v_2, .., v_n)$, the **Euclidean distance** $d_2(u, v)$ is:

$$d_2(u, v) = \|u - v\|_2 = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}$$

The Manhattan distance $d_1(u, v)$ is calculated as:

$$d_1(u, v) = \|u - v\|_1 = \sum_{i=1}^{n} |u_i - v_i|$$




Manhattan distance: $d = |X_2 - X_1| + |Y_2 - Y_1|$

# Inner Products

- Given two vectors $\boldsymbol{x} = (x_1, x_2)$ and $\boldsymbol{y} = (y_1, y_2)$ from the plane, we define their inner product by:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_1 y_1 + x_2 y_2,$$

this can be shown that

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|x\| \|y\| \cos \alpha$$

Where $\alpha$ is the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$

$$\alpha = \cos^{-1} \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|}$$

- We can use the inner products to determine whether two vectors are orthogonal, as this happens if and only if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$ holds.

Let $V$ be a real vector space. The function $\langle \cdot,\cdot \rangle: V \times V \to \mathbb{R}$ is called an inner product if the following holds for all $x, y, z \in V$ and $a \in \mathbb{R}$.

1. $\langle a\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = a\langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$ (linearity of the first variable
2. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ (symmetry)
3. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0$ for all $\boldsymbol{x} \neq 0$ (positive definiteness)

- $\langle 0, \boldsymbol{x} \rangle = \langle 0\boldsymbol{x}, \boldsymbol{x} \rangle = 0\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$
- Special case: $\langle 0, 0 \rangle = 0$
- If $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$, then $\boldsymbol{x} = 0$
- Due to the symmetry and linearity of the first variable, inner products are also linear in second variable. They are called bilinear.
- $\langle \boldsymbol{x}, b\boldsymbol{y} + \boldsymbol{z} \rangle = b\langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$

- The canonical and most prevalent example of inner product spaces in $\mathbb{R}^n$, where the inner product $\langle\cdot,\cdot\rangle$ is defined by:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{y}_i, \qquad \boldsymbol{x} = (x_i, \dots, x_n), \boldsymbol{y} = (y_1, \dots, y_n)$$

This bilinear function is often called the ***dot product.***

# The generated norm

- The 2-norm was defined by $\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$, which, according to our definition of the inner product there, equals $\sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$
- Inner products can be used to define norms on vector spaces.

# Cauchy-Schwarz Inequality

- Let $V$ be an inner product space. Then, for any $x, y \in V$, the inequality $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2 \leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle$ holds.
- The Cauchy-Schwarz inequality is probably one of the most useful tools in studying inner product spaces.

- Let $V$ be an inner product space. Then, the function $\|\cdot\|: V \to [0, \infty)$ defined by $\|x\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ is an norm on $V$.

# Triangle Inequality

- The triangle inequality follows from the Cauchy-Schwarz inequality:

$$
\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y} \rangle \\
&= \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle \\
&\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \\
&\leq (\|x\| + \|y\|)^2
\end{aligned}
$$

$$\|x + y\| \leq \|x\| + \|y\|$$

# Orthogonality

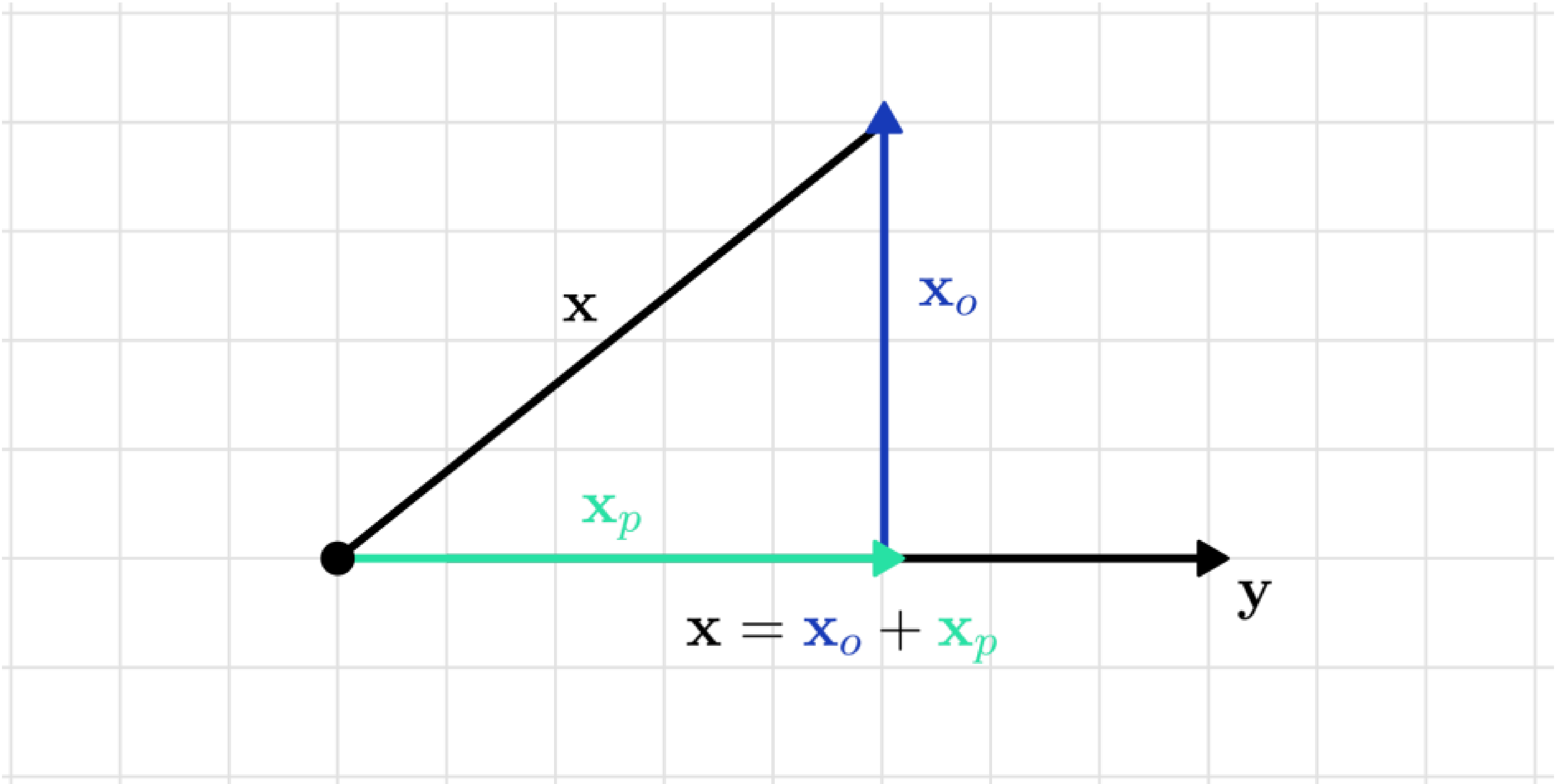
- Let $V$ be an inner product space, and let $x, y \in V$.

  We say that $x$ and $y$ are orthogonal if $\langle x, y \rangle = 0$
- Orthogonality is denoted by as $x \perp y$

- Let $V$ be an inner product space, and let $x, y \in V$. Then, $x$ and $y$ are orthogonal if and only if $\langle x + y, x + y \rangle = \langle x, x \rangle + \langle y, y \rangle$
- This can be expressed as: $\|x + y\|^2 = \|x\|^2 + \|y\|^2$

# Geometric interpretation of inner products

- Using the concept of orthogonality, we can visualize what $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ represent for any $\boldsymbol{x}$ and $\boldsymbol{y}$.




How can we find $x_p$ and $x_o$

- Since $\boldsymbol{x}_p$ has the same direction as $\boldsymbol{y}$, it can be written in the form $\boldsymbol{x}_p = c\boldsymbol{y}$ for some scalar $c \in \mathbb{R}$.
- Because $\boldsymbol{x}_p$ and $\boldsymbol{x}_o$ sum up to $x$, we also have $\boldsymbol{x}_o = \boldsymbol{x} - \boldsymbol{x}_p = \boldsymbol{x} - c\boldsymbol{y}$
- Since $x_o$ is orthogonal to $y$, the constant $c$ can be determined by solving the equation: $\langle \boldsymbol{x} - c\boldsymbol{y}, \boldsymbol{y} \rangle = 0$

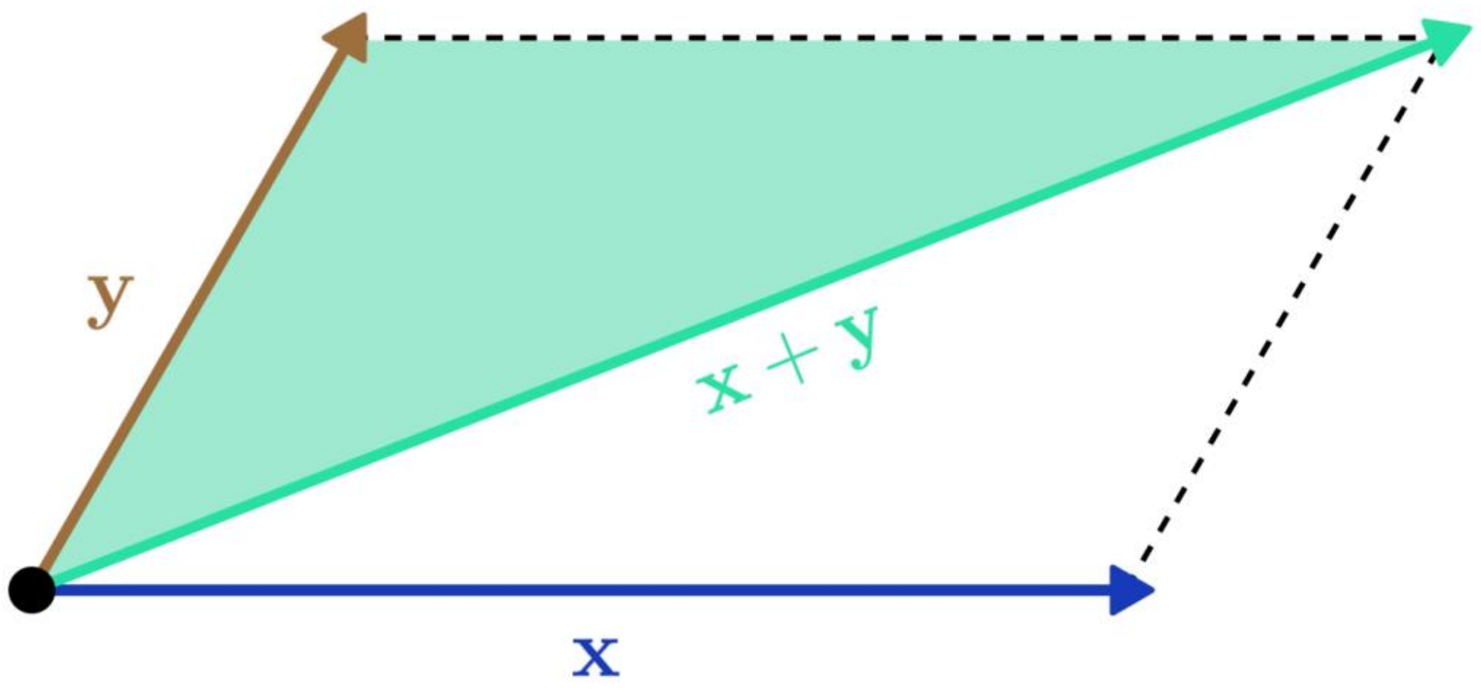
- By using the bilinearity of the inner product, we can express $c$ as: $c = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{y}, \boldsymbol{y} \rangle}$

- So, $x_p = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{y}, \boldsymbol{y} \rangle} \boldsymbol{y}$,

$$x_o = \boldsymbol{x} - \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{y}, \boldsymbol{y} \rangle} \boldsymbol{y}$$

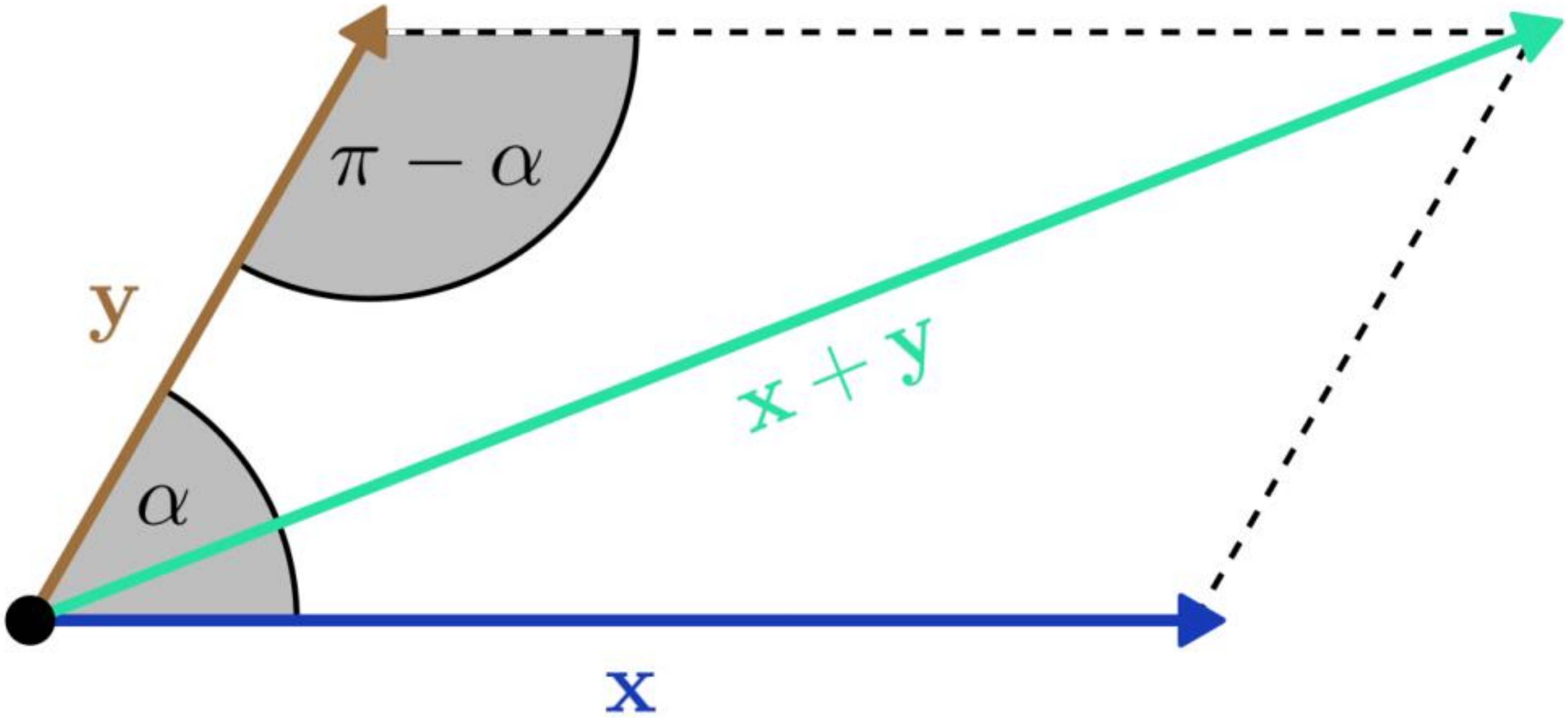
❑We call $x_p$ the *orthogonal projection* of $x$ onto $y$.

$$proj_y(x) = \frac{\langle \boldsymbol{x},\boldsymbol{y} \rangle}{\langle \boldsymbol{y},\boldsymbol{y} \rangle}\boldsymbol{y}$$




- We can use inner product to define the orthogonality relation between two vectors.

$$\langle x+y, x+y\rangle = \|x+y\|^2$$
$$= \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle$$



- Considering that $x, y$ and $x + y$ form a triangle, we can use the law of cosines to express $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y} \rangle = \|\boldsymbol{x} + \boldsymbol{y}\|^2$
- The law of cosines implies:

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\|\boldsymbol{x}\|\|\boldsymbol{y}\| \cos(\pi - \alpha)$$

- Finally, we obtain: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\|\|\boldsymbol{y}\| \cos \alpha$

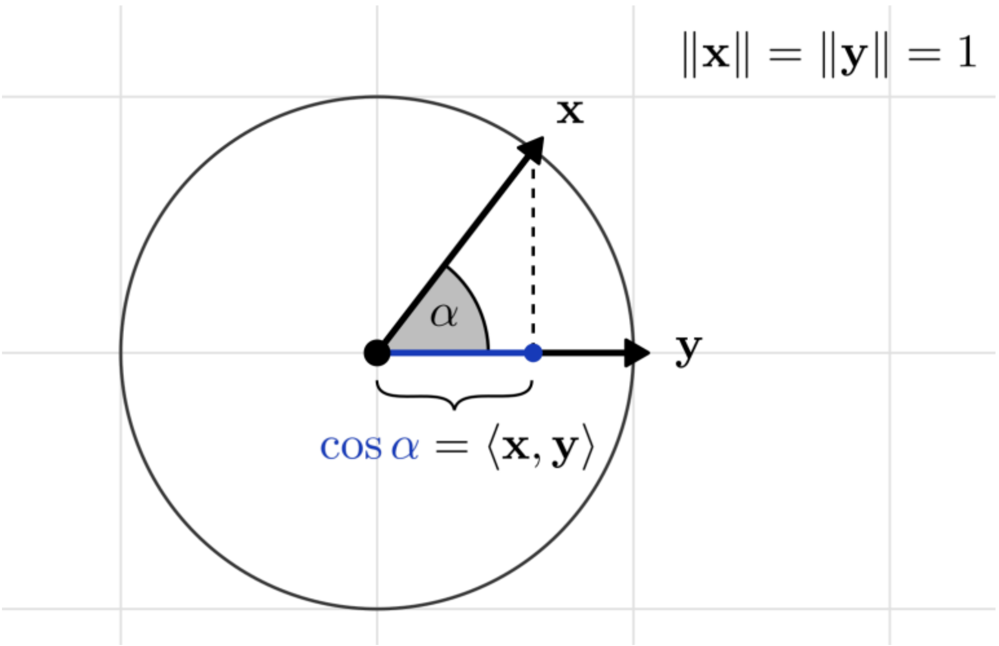
$$\cos\alpha = \frac{\langle x,y\rangle}{\|x\|\|y\|} \rightarrow$$ ***Cosine similarity***

$$\alpha = \cos^{-1}\frac{\langle x, y\rangle}{\|x\|\|y\|}$$

- When both $\boldsymbol{x}$ and $\boldsymbol{y}$ have unit norms, the orthogonal projection equal:

$$proj_y(x) = \langle x, y \rangle y \qquad (\|x\| = \|y\| = 1)$$

- $\langle x, y \rangle$ precisely describes the signed magnitude of the orthogonal projection.
- It can be negative when $proj_y(x)$ and $y$ have an opposite direction.




- ❑Any vector $x$ can be scaled to unit vector norm with the transformation $\boldsymbol{x} \mapsto \frac{\boldsymbol{x}}{||\boldsymbol{x}||}$

- We define the cosine similarity by

$$\cos(x, y) = \left\langle \frac{\boldsymbol{x}}{||\boldsymbol{x}||}, \frac{\boldsymbol{y}}{||\boldsymbol{y}||} \right\rangle$$

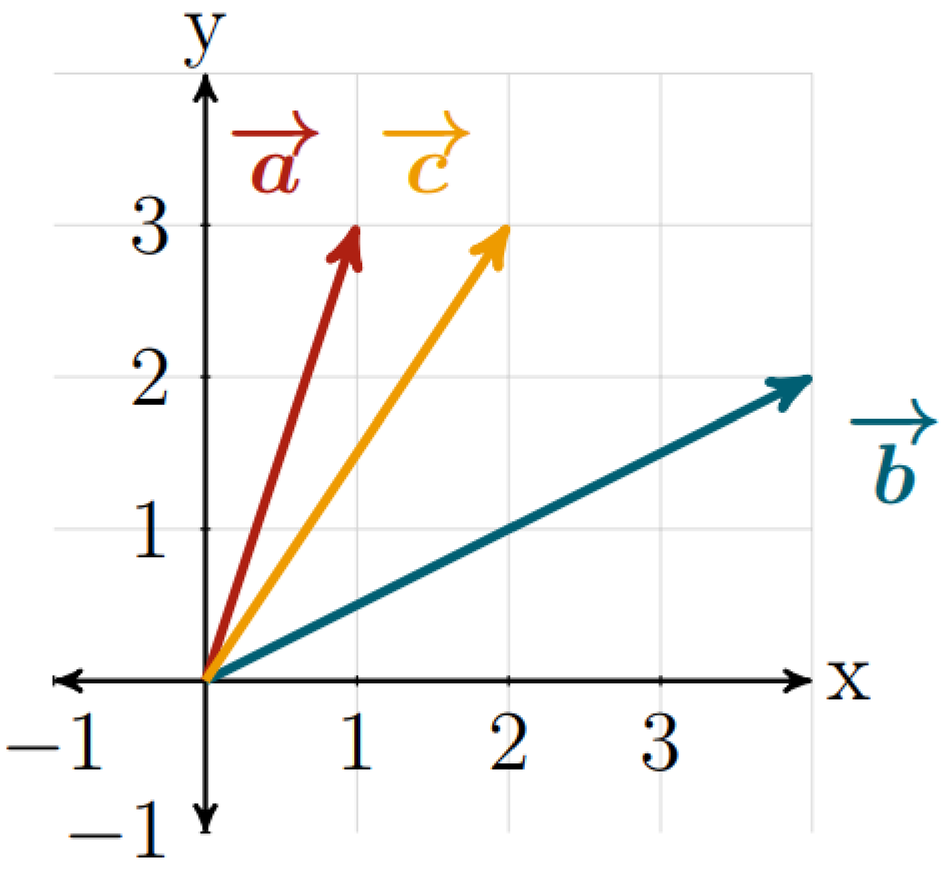
- If $x$ and $y$ represent the feature vectors of two data samples, $\cos(x, y)$ tells us how much the feature move together.

- Because of the scaling, two samples with a high cosine similarity can be far from each other.

# Let’s practice

- Our task is to recommend a movies to a user called, Jessica Martins.
- We know that Susan has watched movie $\vec{a}$, represented by: $\vec{a} = \binom{1}{3}$
- In our library, we have two more movies that we could recommend to Jessica, movies $\vec{b}$ and $\vec{c}$:

$$\vec{b} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \vec{c} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$



$$\cos(a, b) = 0.7071\,, \qquad \cos(a, c) = 0.9648$$

- It's essential to understand the behaviour of the cosine in this context: this function ranges from $-1$ to $1$, where $1$ indicates that the vectors are **perfectly aligned** (*pointing in the same direction*), $0$ shows **orthogonality** (*no similarity*), and $-1$ means that the vectors are opposed.

- In this example, a higher cosine value signifies greater **similarity**.
- Therefore, the movie represented by vector $\vec{c}$, with a cosine of $0.9648$ will be recommend to Jessica as compared to the movie with cosine value of $0.7071$

# Orthogonal and Orthonormal bases

- Through the lens of similarity, orthogonality means that one vector does not contain "***information***" about the other.
- Recall that during the introduction of basis vectors our motivation was to find a minimal set of vectors that can use to express any other vectors.

- Let $V$ be a vector space and $S = \{v_1, \ldots, v_n\}$ is basis. We say that $S$ is an orthogonal basis if $\langle v_i, v_j \rangle = 0$ wherever $i \neq j$.

- Moreover, $S$ is called orthonomal if

$$\langle v_i, v_j \rangle = \begin{cases} 1, & if\ i = j, \\ 0, & if \neq j \end{cases}$$

- $S$ is orthonormal if, in addition to being orthogonal, each vector has unit norm.

- Orthogonal and orthonormal bases are extremely convenient to use.
- If a basis is orthogonal, we can easily obtain an orthonormal basis by simply scaling its vectors to unit norm.

- Let $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ be an arbitrary basis and let $x$ be an arbitrary vector.
- We know that: $\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{v}_i$
- How do we find the coefficients $x_i$?
- If $\{u_i\}_{i=1}^{n}$ is orthonormal, the situation above is much simpler.

- Let $V$ be a vector space and $S = \{u_1, \dots, u\}$ be an orthonormal basis of $V$. Then, for any $x \in V$, $x = \sum_{i=1}^{n} \langle \boldsymbol{x}, \boldsymbol{u}_i \rangle \boldsymbol{u}_i$ holds.

Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k$ be an orthonormal basis for a $k$-dimensional subspace $V \subseteq \mathbb{R}^2$. Then one can write any vector $v \in V$ as a linear combination

$$v = c_1\boldsymbol{u}_1 + \cdots + c_k\boldsymbol{u}_k$$

in which its coordinates

$$c_i = \langle \boldsymbol{u}_i, \boldsymbol{v} \rangle, \ \ i = 1, \dots, k,$$

*are explicitly given as inner products.*

Moreover, its norm is given by the Pythagorean formula $\|v\| = \sqrt{c_1^2 + \cdots + c_k^2} = \sqrt{\sum_{i=1}^{k} \langle u_i, v \rangle^2}$

Using the basis $\binom{1}{1}$ and $\binom{1}{-1}$, find the linear combination of the vector $\binom{3}{2}$

# The Gram-Schmidt Process

- Orthogonal bases are awesome and all, but how do we find them?
- Gram-Schmidt orthogonalization process is used to solve this problem.

- The algorithm takes any set of basis vectors $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ and output an orthonormal basis $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ such that $span(v_1, \dots, v_k) = span(e_1, \dots e_k), k = 1, \dots, n$ that is, the subspace generated by the first $k$ vectors of both sets match.

- Let’s focus on finding an orthogonal system first, which we can normalize later to achieve orthonormality.
- We are going to build our set $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ iteratively.
- It is clear that $\boldsymbol{e}_1 \coloneqq \boldsymbol{v}_1$
- Next thing is to find $\boldsymbol{e}_2$ such that $\boldsymbol{e}_2 \perp \boldsymbol{e}_1$, and together they span the same subspace as $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$

- From geometric interpretation of orthogonality, $\boldsymbol{v}_2$ can be projected onto subspace generated by $\boldsymbol{e}_1$.

$$\boldsymbol{e}_2 := \boldsymbol{v}_2 - proj_{e_1}(\boldsymbol{v}_2) = \boldsymbol{v}_2 - \frac{\langle \boldsymbol{v}_2, \boldsymbol{e}_1 \rangle}{\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle} \boldsymbol{e}_1$$

- It is clear that $\boldsymbol{e}_2 \perp \boldsymbol{e}_1$, and it is also clear that $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ spans the same subspace as $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$

- In the next step, we perform the same process.
- We project $\boldsymbol{v}_3$ onto the subspace generated by $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, and then define $\boldsymbol{e}_3$ as the difference of $\boldsymbol{v}_3$ and the projection.

$$\boldsymbol{e}_3 \coloneqq \boldsymbol{v}_3 - proj_{e_{1,}e_2}(\boldsymbol{v}_3) = \boldsymbol{v}_3 - \frac{\langle \boldsymbol{v}_2, \boldsymbol{e}_1 \rangle}{\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle}\boldsymbol{e}_1 - \frac{\langle \boldsymbol{v}_3, \boldsymbol{e}_2 \rangle}{\langle \boldsymbol{e}_2, \boldsymbol{e}_2 \rangle}\boldsymbol{e}_2$$

- In general, if we have $e_1, \dots, e_k$, the vector $e_{k+1}$ can be found by:

$\boldsymbol{e}_{k+1} \coloneqq \boldsymbol{v}_{k+1} - proj_{\boldsymbol{e}_1,\dots,\boldsymbol{e}_k}(\boldsymbol{v}_{k+1})$, where

$proj_{\boldsymbol{e}_1,\dots,\boldsymbol{e}_k}(x) = \sum_{i=1}^{k} \frac{\langle x, \boldsymbol{e}_i \rangle}{\langle \boldsymbol{e}_i, \boldsymbol{e}_i \rangle} \boldsymbol{e}_i$ is the generalized orthogonal operator, projecting a vector to the subspace generated by $\{e_1, \dots, e_k\}$

# Gram-Schmidt orthogonalization process

- Let $V$ be an inner product vector space and $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\} \subseteq V$ be a set of linearly independent vectors. Then, there exists an orthonormal set $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\} \subseteq V$ such that $span(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k) = span(v_1, \dots, v_k)$ holds for any $k = 1, \dots, n$.

***Inputs:*** *A set of linearly independent vectors* $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\} \subseteq V$.

***Output:*** *A set of orthonormal vectors* $\{e_1, \dots, e_n\}$ *such that*

$$span(\boldsymbol{e}_1, \dots, \boldsymbol{e}_k) = span(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$$

*holds for any* $k = 1, \dots, n$

- What happens if we apply the Gram-Schmidt orthogonalization process to a set of linearly dependent vectors?
- *The algorithm will produce zero (0) in the output.*

Use the Gram-Schmidt process to determine an orthonormal basis for $\mathbb{R}^3$ with the following sets of vectors:

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 3 \\ 6 \\ -6 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}$$


Thank you
Thank you very much