MOODLE IV: ORTHOGONAALITY AND THE GRAM-SCHMIDT PROCEDURE

Orthogonality

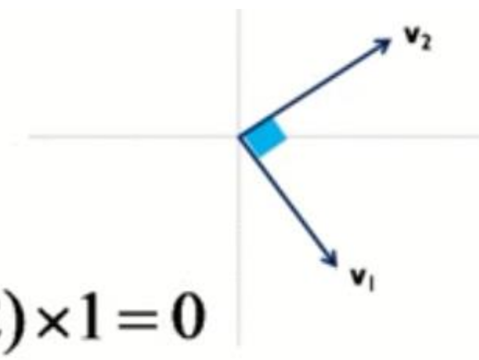
- **What are orthogonal vectors?**
  - The angle between two orthogonal vectors = 90°
  - Also → Dot product = 0

$$\mathbf{v}_1 \bullet \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{v}_2 = 0$$

$$\mathbf{v}_1 \bullet \mathbf{v}_2 = \|\mathbf{v}_1\| . \|\mathbf{v}_2\| \cos\theta$$

$$= \|\mathbf{v}_1\| . \|\mathbf{v}_2\| \cos 90^0 = 0$$

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$



$$\mathbf{v}_1 \bullet \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{v}_2 = 1 \times 2 + (-2) \times 1 = 0$$

- # Orthogonal vs. Orthonormal
  - *Orthogonal = perpendicular vectors "$\theta = 90$"*
  - *Orthonormal = orthogonal + unit vectors*

- **Norm of a vector:**

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \Rightarrow \|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

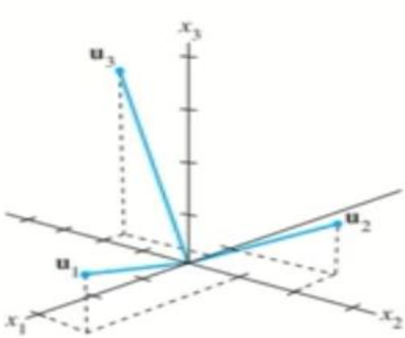
- This is called the "Euclidean norm" or the "ℓ_2-norm"
- There are other types of "norms"

- **Unit vector → Norm = 1 →** $\|\mathbf{u}\| = 1$

- *To get a unit vector u in the same direction as the vector v →* $\dfrac{\mathbf{v}}{\|\mathbf{v}\|}$

- ## Orthogonal Sets
- The set S = $\{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n\}$ is an orthogonal set if

$$\mathbf{v}_i \bullet \mathbf{v}_j = \mathbf{v}_i^T \mathbf{v}_j = 0 \qquad \text{for all } i \neq j$$



$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = 3(-1) + 1(2) + 1(1) = 0$$
$$\mathbf{u}_1 \cdot \mathbf{u}_3 = 3\left(-\tfrac{1}{2}\right) + 1(-2) + 1\left(\tfrac{7}{2}\right) = 0$$
$$\mathbf{u}_2 \cdot \mathbf{u}_3 = -1\left(-\tfrac{1}{2}\right) + 2(-2) + 1\left(\tfrac{7}{2}\right) = 0$$

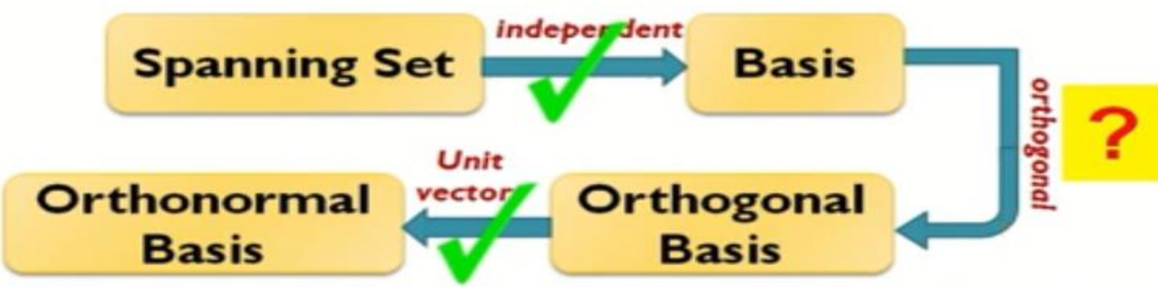
- If we have N vectors, how many dot products do we need to compute to check if they are an orthogonal set?

$$^{N}C_{2}$$

- Example:

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

- **Orthogonal & Orthonormal Basis**
  - **If the vectors in a basis form an orthogonal set, we say that it's an "orthogonal basis"**
  - **If they are ALSO unit vectors, then it's an "orthonormal basis"**



- **Gram-Schmidt Procedure**

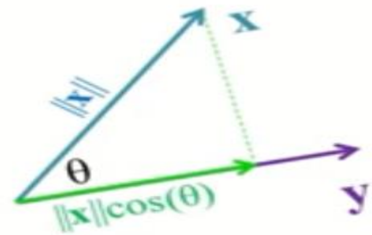
Gram

Schmidt

Laplace

Orthogonal Projection

- We want to project the vector $\mathbf{x}$ onto $\mathbf{y}$

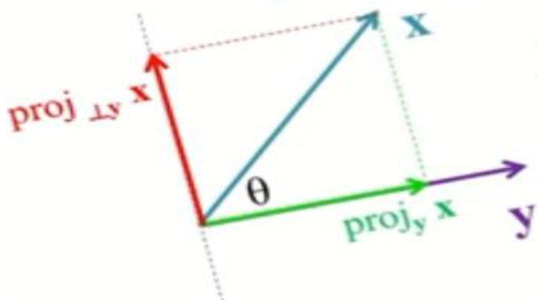


$$\operatorname{proj}_{\mathbf{y}} \mathbf{x} = \left(\|\mathbf{x}\| \cos\theta\right) \frac{\mathbf{y}}{\|\mathbf{y}\|}$$

$$= \left(\cancel{\|\mathbf{x}\|} \frac{\mathbf{x} \bullet \mathbf{y}}{\cancel{\|\mathbf{x}\|}\,\|\mathbf{y}\|}\right) \frac{\mathbf{y}}{\|\mathbf{y}\|}$$

$$= \left(\frac{\mathbf{x} \bullet \mathbf{y}}{\|\mathbf{y}\|^2}\right) \mathbf{y}$$

- We want to project the vector $\mathbf{x}$ in *the direction orthogonal to* $\mathbf{y}$



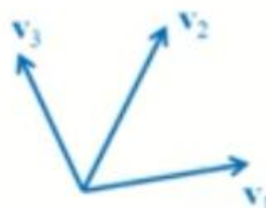
$$\operatorname{proj}_{\perp \mathbf{y}} \mathbf{x} = \mathbf{x} - \operatorname{proj}_{\mathbf{y}} \mathbf{x}$$

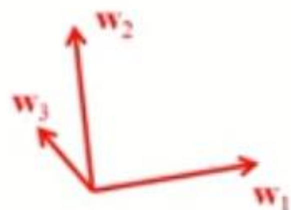
$$= \mathbf{x} - \left(\frac{\mathbf{x} \bullet \mathbf{y}}{\|\mathbf{y}\|^2}\right) \mathbf{y}$$

Gram-Schmidt Procedure

- Applied to a set of independent vectors to get a set of orthogonal vectors that span the same space.

STEPS:





$$\mathbf{w}_1 = \mathbf{v}_1$$

$$\mathbf{w}_2 = \operatorname{proj}_{\perp \mathbf{w}_1} \mathbf{v}_2 = \mathbf{v}_2 - \operatorname{proj}_{\mathbf{w}_1} \mathbf{v}_2$$

$$\mathbf{w}_3 = \operatorname{proj}_{\perp span\{\mathbf{w}_1, \mathbf{w}_2\}} \mathbf{v}_3 = \mathbf{v}_3 - \operatorname{proj}_{\mathbf{w}_1} \mathbf{v}_3 - \operatorname{proj}_{\mathbf{w}_2} \mathbf{v}_3$$

For a set of N independent vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$

$$\mathbf{w}_1 = \mathbf{v}_1$$

$$\mathbf{w}_n = \mathbf{v}_n - \sum_{k=1}^{n-1} \mathrm{proj}_{\mathbf{w}_k} \mathbf{v}_n \qquad \text{for } n = 2,3,\ldots N$$

- **Example:** If $V = \mathrm{span}\{v_1, v_2, v_3\}$, construct an orthonormal basis for V

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

- First: $\{v_1, v_2, v_3\}$ is a spanning set
- Check independence ➔ $\{v_1, v_2, v_3\}$ is an independent set
- Are they orthogonal? NO ➔ $v_1 \bullet v_2 = 3 \neq 0$
- We need to use the Gram-Schmidt Procedure

$$\mathbf{w}_1 = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{w}_2 = \mathrm{proj}_{\perp \mathbf{w}_1} \mathbf{v}_2 = \mathbf{v}_2 - \left(\frac{\mathbf{v}_2 \bullet \mathbf{w}_1}{\|\mathbf{w}_1\|^2} \right) \mathbf{w}_1$$

$$= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} \Rightarrow \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{w}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{w}_3 = \operatorname{proj}_{\perp span\{\mathbf{w}_1, \mathbf{w}_2\}} \mathbf{v}_3 = \mathbf{v}_3 - \left(\frac{\mathbf{v}_3 \bullet \mathbf{w}_1}{\|\mathbf{w}_1\|^2} \right) \mathbf{w}_1 - \left(\frac{\mathbf{v}_3 \bullet \mathbf{w}_2}{\|\mathbf{w}_2\|^2} \right) \mathbf{w}_2$$

$$= \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{12} \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

$$\boxed{\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{w}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{w}_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}} \Rightarrow$$

Orthogonal Basis for V

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \longrightarrow \mathbf{u}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} = \frac{1}{\sqrt{1^2 + 1^2 + 1^2 + 1^2}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{4} \\ 1/\sqrt{4} \\ 1/\sqrt{4} \\ 1/\sqrt{4} \end{bmatrix}$$

$$\mathbf{w}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} \longrightarrow \mathbf{u}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \frac{1}{\sqrt{(-3)^2 + 1^2 + 1^2 + 1^2}} \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{12}} \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \end{bmatrix}$$

$$\mathbf{w}_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix} \longrightarrow \mathbf{u}_3 = \frac{\mathbf{w}_3}{\|\mathbf{w}_3\|} = \frac{1}{\sqrt{0^2+(-2)^2+1^2+1^2}} \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$$

$$\boxed{\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{4} \\ 1/\sqrt{4} \\ 1/\sqrt{4} \\ 1/\sqrt{4} \end{bmatrix} \quad \mathbf{u}_2 = \begin{bmatrix} -3/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \end{bmatrix} \quad \mathbf{u}_3 = \begin{bmatrix} 0 \\ -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}}$$

$\Rightarrow$ Orthonormal Basis for V

Review Questions

- ***Which of the following statements are true?***
  - Any orthogonal set is an independent set
  - A vector space has only one orthonormal basis
  - Any set of orthogonal vectors in $\mathbb{R}^3$ is a basis for $\mathbb{R}^3$
  - If $\{v_1, v_2, v_3\}$ is a basis for V, and x is orthogonal to each of v_1, v_2 and v_3, then x is orthogonal to all vectors in V

Any vector in V $\rightarrow$ $w = c_1v_1 + c_2v_2 + c_3v_3$

$x \bullet w = x \bullet (c_1v_1 + c_2v_2 + c_3v_3)$

$= c_1(x \bullet v_1) + c_2(x \bullet v_2) + c_3(x \bullet v_3) = 0$

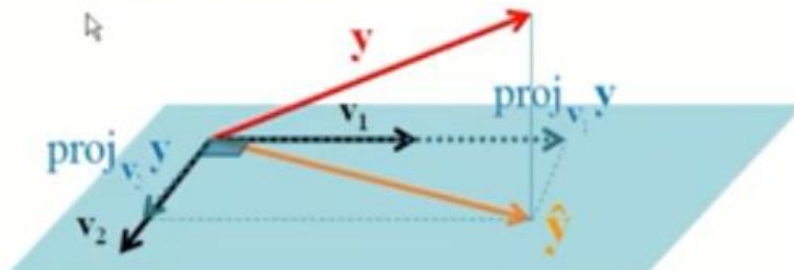
Projection of a vector onto a vector space

- The orthogonal projection of a vector $\mathbf{y}$ onto a vector space W is

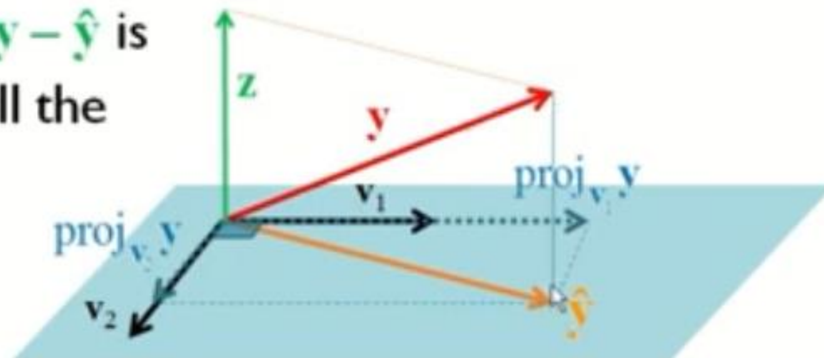
$$\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$$

- If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthogonal basis of W, then

$$\hat{\mathbf{y}} = \sum_{i=1}^{n} \text{proj}_{\mathbf{v}_i} \mathbf{y}$$



- The vector $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to all the vectors in W.



Projection of a vector onto a vector space

- The set of all vectors orthogonal to the vector space W is called $W^{\perp}$ (W perpendicular or "W *perp*")
- $W^{\perp}$ is called the orthogonal complement of W
- $\mathbf{z} \in W^{\perp}$

Orthogonal Matrix

- **What is an orthogonal matrix?**
  - A square matrix whose columns are *orthonormal*

$$\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$$

  - Let $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3]$ where $\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3$ are the columns of $\mathbf{A}$

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \mathbf{a}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T\mathbf{a}_1 & \mathbf{a}_1^T\mathbf{a}_2 & \mathbf{a}_1^T\mathbf{a}_3 \\ \mathbf{a}_2^T\mathbf{a}_1 & \mathbf{a}_2^T\mathbf{a}_2 & \mathbf{a}_2^T\mathbf{a}_3 \\ \mathbf{a}_3^T\mathbf{a}_1 & \mathbf{a}_3^T\mathbf{a}_2 & \mathbf{a}_3^T\mathbf{a}_3 \end{bmatrix}$$

$$= \begin{bmatrix} \|\mathbf{a}_1\|^2 & \mathbf{a}_1^T\mathbf{a}_2 & \mathbf{a}_1^T\mathbf{a}_3 \\ \mathbf{a}_2^T\mathbf{a}_1 & \|\mathbf{a}_2\|^2 & \mathbf{a}_2^T\mathbf{a}_3 \\ \mathbf{a}_3^T\mathbf{a}_1 & \mathbf{a}_3^T\mathbf{a}_2 & \|\mathbf{a}_3\|^2 \end{bmatrix}$$

- **Example:** Show that the given matrix is orthogonal and find its inverse

$$\mathbf{A}=\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \end{bmatrix}$$

- **Example:**

$$\mathbf{A}^T\mathbf{A}=\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \end{bmatrix}$$

$$=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}=\mathbf{I}$$

$$\mathbf{A}^T\mathbf{A} = \mathbf{I}$$

$$\mathbf{A}^T = \mathbf{A}^{-1}$$

$$\mathbf{A}^{-1} = \mathbf{A}^T = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$$

- **NOTE:**

$$\mathbf{A} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \end{bmatrix} \longrightarrow$$

Orthogonal matrix because the columns are orthonormal

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 0 \\ 1 & 1 & 1 \end{bmatrix} \longrightarrow$$

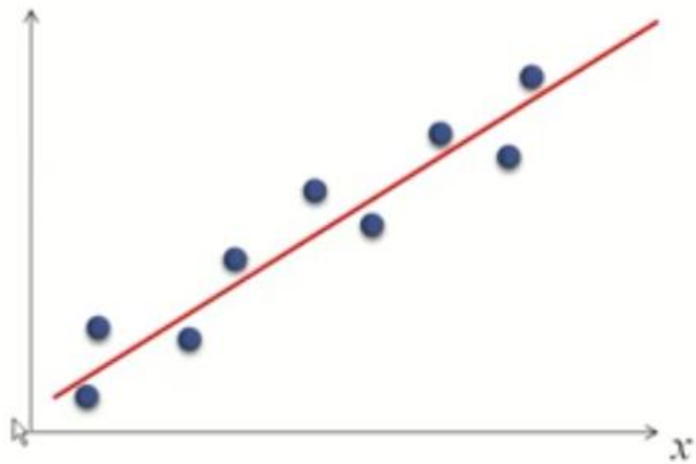
NOT orthogonal The columns are only orthogonal so $\mathbf{B}^T\mathbf{B}$ is a diagonal matrix, but not $\mathbf{I}$

The Least Squares Problem

$$\boxed{y = a\,x + b}$$

$$y_1 = a\,x_1 + b$$
$$y_2 = a\,x_2 + b$$
$$y_3 = a\,x_3 + b$$
$$\vdots$$
$$y_n = a\,x_n + b$$

n equations in two unknowns a and b with NO solution



- Given an inconsistent system of linear equations

$$\mathbf{Ax} = \mathbf{b}$$

that has no solution.

- We still need a solution
- There is no $\mathbf{x}$ that satisfies $\mathbf{Ax} = \mathbf{b}$, so we try to find a vector $\mathbf{x}$ that makes $\mathbf{Ax}$ *as close as possible to* $\mathbf{b}$

As close as possible → Minimum distance → $\min\|\mathbf{b} - \mathbf{Ax}\|$ → Projection

- Recall that if $\mathbf{Ax} = \mathbf{v}$ is consistent then $\mathbf{v}$ is a linear combination of the columns of $\mathbf{A}$

$$\mathbf{v} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n$$

- All the $\mathbf{v}$'s that make $\mathbf{Ax} = \mathbf{v}$ a consistent system is a subspace called ***the column space of* $\mathbf{A}$** because it is spanned by the columns of $\mathbf{A}$
- Since $\mathbf{Ax} = \mathbf{b}$ is inconsistent, then $\mathbf{b}$ is NOT a linear combination of the columns of $\mathbf{A}$
- The closest $\mathbf{v}$ to the vector $\mathbf{b}$ is the one we're looking for, in order to find $\hat{\mathbf{x}}$
- This vector $\mathbf{v}$ lies in the column space of $\mathbf{A}$

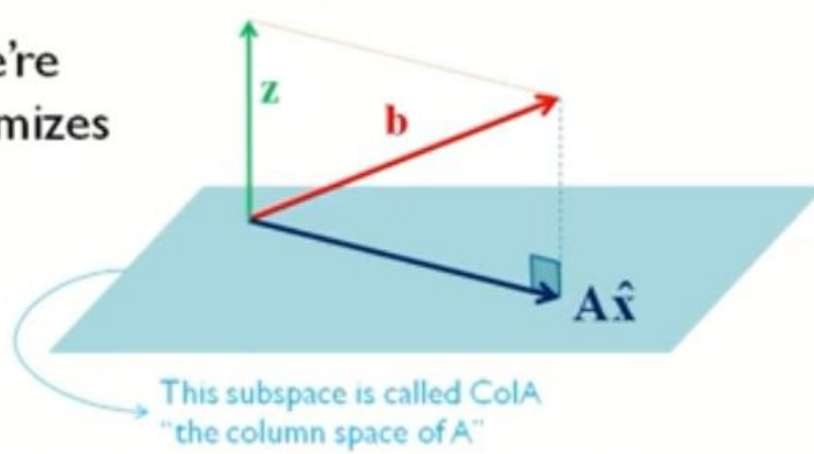
The Least Squares Problem

- We want to project $\mathbf{b}$ onto the subspace containing all possible vectors of the form $\mathbf{Ax}$ (i.e. all the $\mathbf{v}$'s that make $\mathbf{Ax} = \mathbf{v}$ a consistent system) ➔ $\text{Col}\mathbf{A}$

The vector $\hat{\mathbf{x}}$ we're looking for minimizes $\|\mathbf{b} - \mathbf{Ax}\|$

$$\mathbf{b} = \mathbf{z} + \mathbf{A}\hat{\mathbf{x}}$$
$$\mathbf{z} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$$

Find $\hat{\mathbf{x}}$



If $\mathbf{z}$ is orthogonal to all the vectors $\mathbf{v}$ in the column space, then it is orthogonal to the columns of $\mathbf{A}$

$$\mathbf{A}^T\mathbf{z} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{a}_1^T\mathbf{z} \\ \mathbf{a}_2^T\mathbf{z} \\ \vdots \\ \mathbf{a}_n^T\mathbf{z} \end{bmatrix} = \mathbf{0}$$

$$\mathbf{A}^T\mathbf{z} = \mathbf{0}$$

$$\mathbf{A}^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0}$$

$$\mathbf{A}^T\mathbf{b} - \mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{0}$$

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{b}$$

$$\hat{\mathbf{x}} = \boxed{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T}\mathbf{b}$$

Pseudo-inverse of $\mathbf{A}$

- **Example 1:** Find the least squares solution of the inconsistent system $\mathbf{A}\mathbf{x} = \mathbf{b}$

$$\mathbf{A} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

and determine the least squares error

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$\mathbf{A}^T\mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{b}$$

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

- The least squares error is $\|\mathbf{b}-\mathbf{A}\mathbf{x}\|$

$$\mathbf{b}-\mathbf{A}\hat{\mathbf{x}}=\begin{bmatrix}2\\0\\11\end{bmatrix}-\begin{bmatrix}4&0\\0&2\\1&1\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}2\\0\\11\end{bmatrix}-\begin{bmatrix}4\\4\\3\end{bmatrix}=\begin{bmatrix}-2\\-4\\8\end{bmatrix}$$

$$\|\mathbf{b}-\mathbf{A}\hat{\mathbf{x}}\|=\sqrt{(-2)^2+(-4)^2+8^2}=9.165$$

- **Example 2:** Find the least squares solution of the inconsistent system $\mathbf{A}\mathbf{x}=\mathbf{b}$

$$\mathbf{A}=\begin{bmatrix}1&1&0&0\\1&1&0&0\\1&0&1&0\\1&0&1&0\\1&0&0&1\\1&0&0&1\end{bmatrix},\quad \mathbf{b}=\begin{bmatrix}-3\\-1\\0\\2\\5\\1\end{bmatrix}$$

Solution

$$\mathbf{A}=\underset{6\times 4}{\begin{bmatrix}1&1&0&0\\1&1&0&0\\1&0&1&0\\1&0&1&0\\1&0&0&1\\1&0&0&1\end{bmatrix}},\quad \mathbf{b}=\underset{6\times 1}{\begin{bmatrix}-3\\-1\\0\\2\\5\\1\end{bmatrix}}$$

$$\mathbf{A}^T\mathbf{A} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}}_{4\times6} \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}}_{6\times4} = \underbrace{\begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}}_{4\times4}$$

$$\mathbf{A}^T\mathbf{b} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}}_{4\times6} \underbrace{\begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}}_{6\times1} = \underbrace{\begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}}_{4\times1}$$

Example 2:

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$\mathbf{A}^T\mathbf{b} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{b}$$

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

After several steps of elimination

$$\sim \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} \boxed{1} & 0 & 0 & 1 & 3 \\ 0 & \boxed{1} & 0 & -1 & -5 \\ 0 & 0 & \boxed{1} & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \qquad \begin{aligned} x_1 &= 3 - x_4 \\ x_2 &= -5 + x_4 \\ x_3 &= -2 + x_4 \end{aligned}$$

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

So there's more than one $\hat{\mathbf{x}}$ that minimizes $||b - A\hat{\mathbf{x}}||$

INNER PRODUCT SPACES

- The conditions for a set to be called a **vector space** involve ***addition*** and ***scalar multiplication***

- If we define a ***product operation*** with certain conditions we get an **inner product space**

- This product operation is called an ***inner product***
- The ***dot product*** is an inner product, but not the only inner product there is!

What is an inner product?

- It is an operation that associates each pair of vectors in a vector space V to a scalar quantity known as the inner product of the two vectors.
- *It satisfies the following axioms:*
  - For every $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ where V is a vector space, the inner product $<\mathbf{u}, \mathbf{v}>$ satisfies:

1. $<\mathbf{u}, \mathbf{v}> = <\mathbf{v}, \mathbf{u}>$ → Commutative
2. $<\mathbf{u} + \mathbf{v}, \mathbf{w}> = <\mathbf{u}, \mathbf{w}> + <\mathbf{v}, \mathbf{w}>$ → Distributive
3. $<c\mathbf{u}, \mathbf{v}> = c<\mathbf{u}, \mathbf{v}>$ where $c \in \mathbb{R}$
4. $<\mathbf{u}, \mathbf{u}> \geq 0$ and $<\mathbf{u}, \mathbf{u}> = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Inner Product Space:

- An inner product space is a vector space on which an inner product operation is defined.

Vector Space

1. Addition
2. Scalar Multiplication

Inner Product Space

1. Addition
2. Scalar Multiplication
3. Inner Product

- **Example:**
  - Which of these represents an inner product?

1. The dot product $\mathbf{u} \bullet \mathbf{v} = \mathbf{u}^T\mathbf{v}$ on $\mathbb{R}^2$

Answer:

1 $\mathbf{u}^T\mathbf{v} = u_1v_1 + u_2v_2$

$\mathbf{v}^T\mathbf{u} = v_1u_1 + v_2u_2 = u_1v_1 + u_2v_2 = \mathbf{u}^T\mathbf{v}$

2 $(\mathbf{u}+\mathbf{w}) \bullet \mathbf{v} = (\mathbf{u}+\mathbf{w})^T\mathbf{v} = (\mathbf{u}^T + \mathbf{w}^T)\mathbf{v} = \mathbf{u}^T\mathbf{v} + \mathbf{w}^T\mathbf{v}$

$\mathbf{u} \bullet \mathbf{v} + \mathbf{w} \bullet \mathbf{v} = \mathbf{u}^T\mathbf{v} + \mathbf{w}^T\mathbf{v}$

3 $(c\mathbf{u})^T\mathbf{v} = cu_1v_1 + cu_2v_2$

$c(\mathbf{u}^T\mathbf{v}) = c(u_1v_1 + u_2v_2) = cu_1v_1 + cu_2v_2 = (c\mathbf{u})^T\mathbf{v}$

4 If $\mathbf{u} = \mathbf{0} \Rightarrow \mathbf{u} \bullet \mathbf{u} = \mathbf{0}^T\mathbf{0} = 0$ (or $0^2 + 0^2 = 0$)

If $\mathbf{u} \bullet \mathbf{u} = 0 \Rightarrow u_1^2 + u_2^2 = 0 \Rightarrow u_1 = u_2 = 0 \Rightarrow \mathbf{u} = \mathbf{0}$

2. Matrix multiplication $< \mathbf{A}, \mathbf{B} > = \mathbf{AB}$ on $\mathbb{M}_{2\times 2}$

Answer:

It is NOT an inner product because the commutative property does not hold for matrix multiplication, because $\mathbf{AB} \neq \mathbf{BA}$

3. The product $<\mathbf{u}, \mathbf{v}> = 2u_1v_1 + 3u_2v_2$ on $\mathbb{R}^3$

Answer:

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

1

$$<\mathbf{u}, \mathbf{v}> = 2u_1v_1 + 3u_2v_2$$
$$<\mathbf{v}, \mathbf{u}> = 2v_1u_1 + 3v_2u_2$$
$$<\mathbf{u}, \mathbf{v}> = <\mathbf{v}, \mathbf{u}>$$

2

$$<\mathbf{u} + \mathbf{v}, \mathbf{w}>$$
$$= 2(u_1 + v_1)w_1 + 3(u_2 + v_2)w_2$$
$$= 2u_1w_1 + 2v_1w_1 + 3u_2w_2 + 3v_2w_2$$
$$= (2u_1w_1 + 3u_2w_2) + (2v_1w_1 + 3v_2w_2) = <\mathbf{u}, \mathbf{w}> + <\mathbf{v}, \mathbf{w}>$$

3

$$<c\mathbf{u}, \mathbf{v}>$$
$$= 2(cu_1)v_1 + 3(cu_2)v_2$$
$$= c(2u_1v_1 + 3u_2v_2)$$
$$= c<\mathbf{u}, \mathbf{v}>$$

4

→ If $\mathbf{u} = \mathbf{0}$

$$<\mathbf{u}, \mathbf{u}> = 2u_1u_1 + 3u_2u_2 = 2u_1^2 + 3u_2^2 = 2(0)^2 + 3(0)^2 = 0$$

→ If $<\mathbf{u}, \mathbf{u}> = 0$

$2u_1{}^2 + 3u_2{}^2 = 0$ so $u_1 = u_2 = 0$

But what about u_3 ? → It doesn't have to be zero!

$$\left\langle \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} \right\rangle = 2(0)(0) + 3(0)(0) = 0$$

Here $\mathbf{u} \neq \mathbf{0}$ but $<\mathbf{u},\mathbf{u}> = 0$ ➔ Not an inner product

If it were defined on $\mathbb{R}^2$, the same definition would be an inner product.

➔ If $<\mathbf{u},\mathbf{u}> = 0$

$2u_1{}^2 + 3u_2{}^2 = 0$ so $u_1 = u_2 = 0 \rightarrow \mathbf{u} = \mathbf{0}$

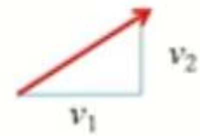
NORM OF A VECTOR

- **Length (Norm) of a Vector**
  - A function $||\mathbf{v}||$ of a vector $\mathbf{v}$ that gives a scalar value is classified as a norm (length) if it satisfies the following conditions:

    1. $||\mathbf{v}|| \geq 0$
    2. $||c\mathbf{v}|| = |c|.||\mathbf{v}||$ where $c \in \mathbb{R}$
    3. $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$ ➔ Triangle Inequality
    4. $||\mathbf{v}|| = 0$ if and only if $\mathbf{v} = \mathbf{0}$

- **Example 1:** The norm of a vector in $\mathbb{R}^2$ defined as

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2}$$



is called the "Euclidean norm" or the ℓ_2-norm.

It is an inner product since it satisfies all the conditions we mentioned

> The Euclidean norm for vectors in $\mathbb{R}^n$
> $$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

➔ $\|\mathbf{v}\| \geq 0$ ✓

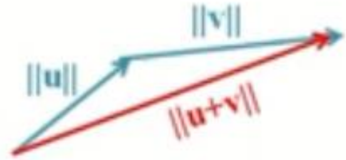
➔ $\|c\mathbf{v}\| = |c|.\|\mathbf{v}\|$ where $c \in \mathbb{R}$

$$c\mathbf{v} = \begin{bmatrix} cv_1 \\ cv_2 \end{bmatrix} \qquad \|c\mathbf{v}\| = \sqrt{(cv_1)^2 + (cv_2)^2} = \sqrt{c^2(v_1^2 + v_2^2)}$$

$$= |c|\sqrt{v_1^2 + v_2^2} = |c|\|\mathbf{v}\|$$

➔ $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$

$$\sqrt{(u_1 + v_1)^2 + (u_2 + v_2)^2} \leq \sqrt{(u_1^2 + u_2^2)} + \sqrt{(v_1^2 + v_2^2)}$$



The sum of lengths of any 2 sides of a triangle is ≥ the length of the 3rd side ✓

➔ $||\mathbf{v}|| = 0$ if and only if $\mathbf{v} = \mathbf{0}$

If $\mathbf{v}$ is the zero vector ($\mathbf{v}=\mathbf{0}$) then obviously

$\|\mathbf{0}\| = \sqrt{0^2 + 0^2} = 0$

If $||\mathbf{v}|| = 0$, then does this mean that $\mathbf{v}$ must be the zero vector?

$\sqrt{v_1^2 + v_2^2} = 0 \Rightarrow v_1^2 + v_2^2 = 0 \Rightarrow \therefore v_1 = v_2 = 0$ ✓

The ℓ_p norm

$$\|x\|_p = \left(\sum_{i=1}^{N} |x_i|^p \right)^{1/p}$$

- For $p \geq 1$, the ℓ_p norm satisfies the norm conditions
- **Examples:**
  - The ℓ_1 norm in $\mathbb{R}^n$ is $||\mathbf{v}|| = |v_1| + |v_2| + \ldots + |v_n|$
  - The ℓ_2 norm in $\mathbb{R}^n$ is the Euclidean norm

Example 2:

The ℓ_1 norm in $\mathbb{R}^n$ is $||\mathbf{v}||_1 = |v_1| + |v_2| + \ldots + |v_n|$.
Verify that it satisfies the conditions of a norm.

1. $||\mathbf{v}|| \geq 0$
2. $||c\mathbf{v}|| = |c|.||\mathbf{v}||$ where $c \in \mathbb{R}$
3. $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$
4. $||\mathbf{v}|| = 0$ if and only if $\mathbf{v} = \mathbf{0}$

→ $\|\mathbf{v}\| \geq 0$

$\|\mathbf{v}\|_1 = |v_1| + |v_2| + \ldots + |v_n| \geq 0$ ***because each*** $|v_i| \geq 0$

→ $\|c\mathbf{v}\| = |c|.\|\mathbf{v}\|$

$$\begin{aligned}\|c\mathbf{v}\|_1 &= |cv_1| + |cv_2| + \ldots + |cv_n| \\ &= |c|.|v_1| + |c|.|v_2| + \ldots + |c|.|v_n| \\ &= |c|.(|v_1| + |v_2| + \ldots + |v_n|) = |c|.\|\mathbf{v}\|\end{aligned}$$

→ $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$

$\|\mathbf{u}+\mathbf{v}\|_1 = |u_1+v_1| + |u_2+v_2| + \ldots + |u_n+v_n|$

$$\begin{aligned}\|\mathbf{u}\|_1+\|\mathbf{v}\|_1 &= (|u_1| + |u_2| + \ldots + |u_n|) \\ &\quad + (|v_1| + |v_2| + \ldots + |v_n|) \\ &= |u_1|+|v_1| + |u_2|+|v_2| + \ldots + |u_n|+|v_n|\end{aligned}$$

→ $|u_i + v_i| \leq |u_i|+|v_i|$

→ $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$

→ If $\mathbf{v} = \mathbf{0}$ then $\|\mathbf{v}\|_1 = |0| + |0| + \ldots = 0$

→ If $\|\mathbf{v}\|_1 = 0$ then $|v_1| + |v_2| + \ldots + |v_n| = 0$

Since $|v_i| \geq 0$ then $v_1 = v_2 = \ldots = v_n = 0$

Example 3:

The ℓ_0 norm in $\mathbb{R}^n$ is $\|\mathbf{v}\|_0 = |v_1|^0 + |v_2|^0 + \dots + |v_n|^0$ for $v_i \neq 0$, which is the number of non-zero elements in the vector $\mathbf{v}$.

Is it a "norm"? (Does it satisfy the conditions?)

- NO
- Which condition(s) does it NOT satisfy?

1. $\|\mathbf{v}\| \geq 0$ ✓ You're counting the number of non-zero elements so it's always ≥ 0
2. $\|c\mathbf{v}\| = |c|.\|\mathbf{v}\|$ where $c \in \mathbb{R}$

$$LHS = |cv_1|^0 + |cv_2|^0 + \dots + |cv_n|^0 = \cancel{|c|^0}^{\,1}(|v_1|^0 + |v_2|^0 + \dots + |v_n|^0)$$

= number of non-zero elements

$$RHS = |c|.\|\mathbf{v}\| = |c|.(|v_1|^0 + |v_2|^0 + \dots + |v_n|^0)$$ ✗

= number of non-zero elements × |c|

! Despite the ℓ_0 norm not being an actual "norm", it is important, since it defines the vector's ***sparsity***

Example: If $\mathbf{x} = [1\ 0\ 0\ 2\ 0\ 3\ 0\ 0]^T$ we say that $\mathbf{x}$ is a 3-sparse vector ($\|\mathbf{x}\|_0 = 3$)

DISTANCE BETWEEN TWO VECTORS

- **Distance between two vectors:**
- Based on how we define the length of a vector, we can define the distance between two vectors.
- Like the norm (length), the function $d(\mathbf{u},\mathbf{v})$ can be called the "distance" between $\mathbf{u}$ and $\mathbf{v}$ if:

1. $d(\mathbf{u},\mathbf{v}) \geq 0$
2. $d(\mathbf{u},\mathbf{v}) = d(\mathbf{v},\mathbf{u})$
3. $d(\mathbf{u},\mathbf{v}) \leq d(\mathbf{u},\mathbf{w}) + d(\mathbf{w},\mathbf{v})$
4. $d(\mathbf{u},\mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$

- **Example:** The Euclidean distance

$$d(\mathbf{u},\mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_2 = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$$

satisfies the conditions for a "distance"

1. $d(\mathbf{u},\mathbf{v}) \geq 0$ ✓
2. $d(\mathbf{u},\mathbf{v}) = d(\mathbf{v},\mathbf{u})$ ✓ $\|\mathbf{u} - \mathbf{v}\|_2 = \|\mathbf{v} - \mathbf{u}\|_2$
3. $d(\mathbf{u},\mathbf{v}) \leq d(\mathbf{u},\mathbf{w}) + d(\mathbf{w},\mathbf{v})$ ✓
4. $d(\mathbf{u},\mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$ ✓