THE UNIVERSITY OF CHICAGO

# MSCA 37016
# Advanced Linear Algebra for Machine Learning

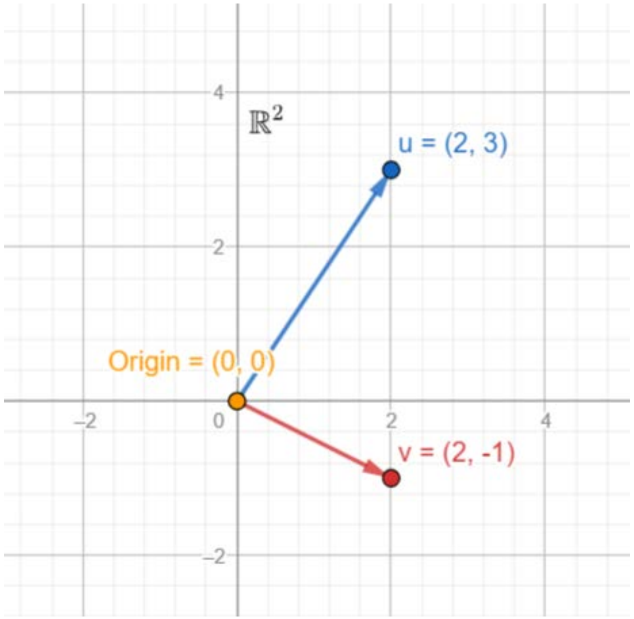
Lecture 1

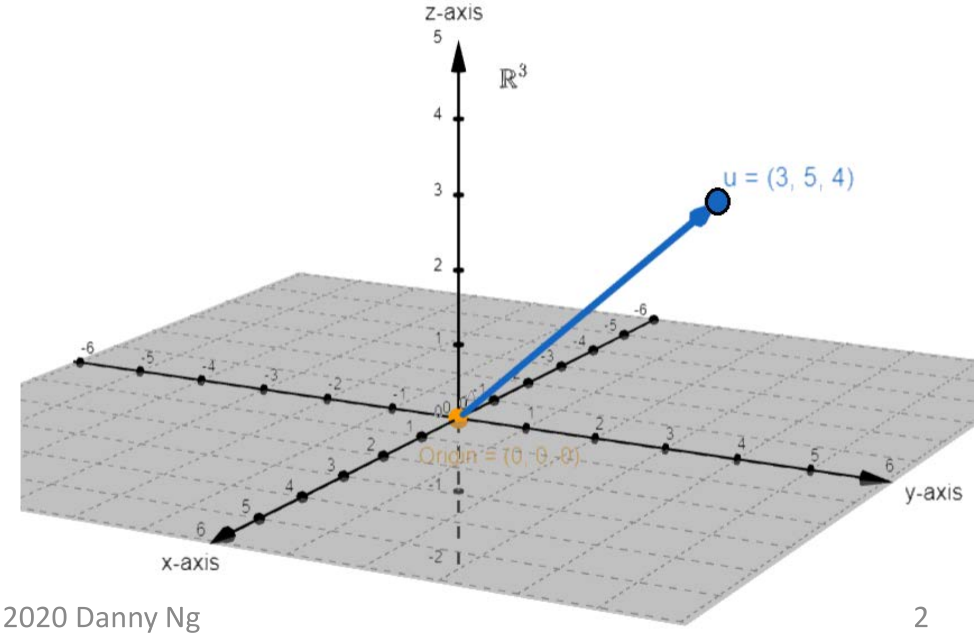
Danny Ng

# Vectors in Euclidean Space $\mathbb{R}^d$: Point, Arrow, and Tuple

- 3 equivalent views of vector
  1. Point: a location in $\mathbb{R}^d$ space
  2. Arrow from origin: magnitude and direction
  3. d-tuple coordinates: $\vec{v} = (v_1, \dots, v_d) \in \mathbb{R}^d$

Zero vector / origin: $\vec{0} = (0, \dots, 0)$

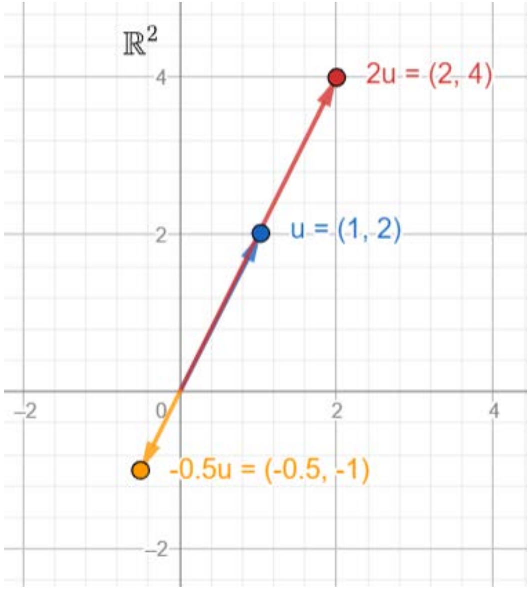


# Vector Scaling, Addition, and Subtraction

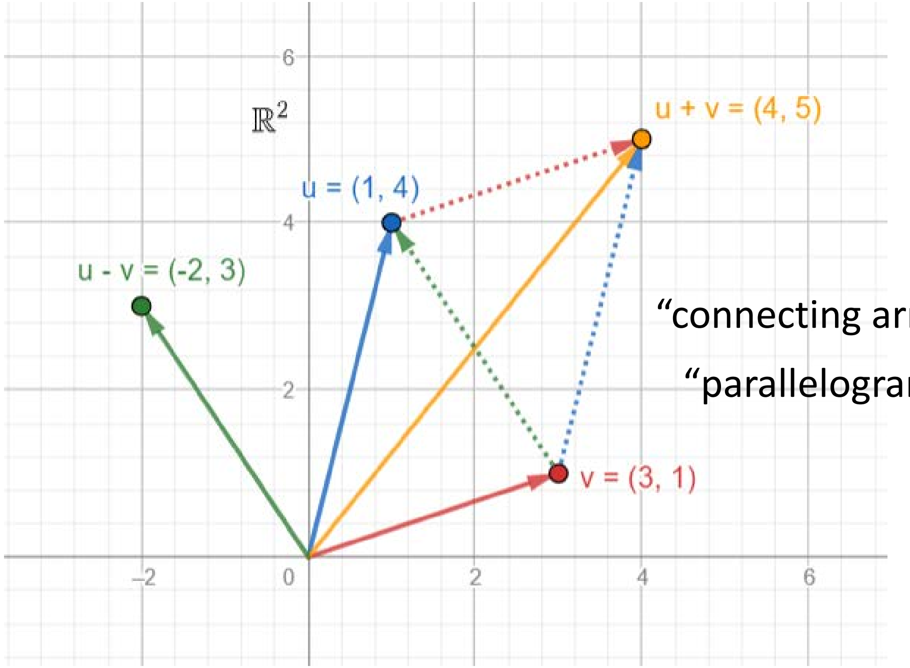
- Vectors $\vec{u}, \vec{v} \in \mathbb{R}^d$ and scalar $a \in \mathbb{R}$

**Scaling:** $a\,\vec{v} = (a\,v_1, \ldots, a\,v_d)$




**Addition:** $\vec{u} + \vec{v} = (u_1 + v_1, \ldots, u_d + v_d)$

**Subtraction:** $\vec{u} - \vec{v} = (u_1 - v_1, \ldots, u_d - v_d)$




"connecting arrows"

"parallelogram"

# Vector Arithmetic Properties

- Commutative

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

- Associative

$$(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$$

- Distributive

$$a(\vec{u} + \vec{v}) = (a\vec{u}) + (a\vec{v})$$

$$(a + b)\vec{u} = (a\vec{u}) + (b\vec{u})$$

e.g. $$3\left(\begin{pmatrix}2\\4\end{pmatrix} + \begin{pmatrix}-2\\1\end{pmatrix}\right) = \begin{pmatrix}0\\15\end{pmatrix} = 3\begin{pmatrix}2\\4\end{pmatrix} + 3\begin{pmatrix}-2\\1\end{pmatrix}$$

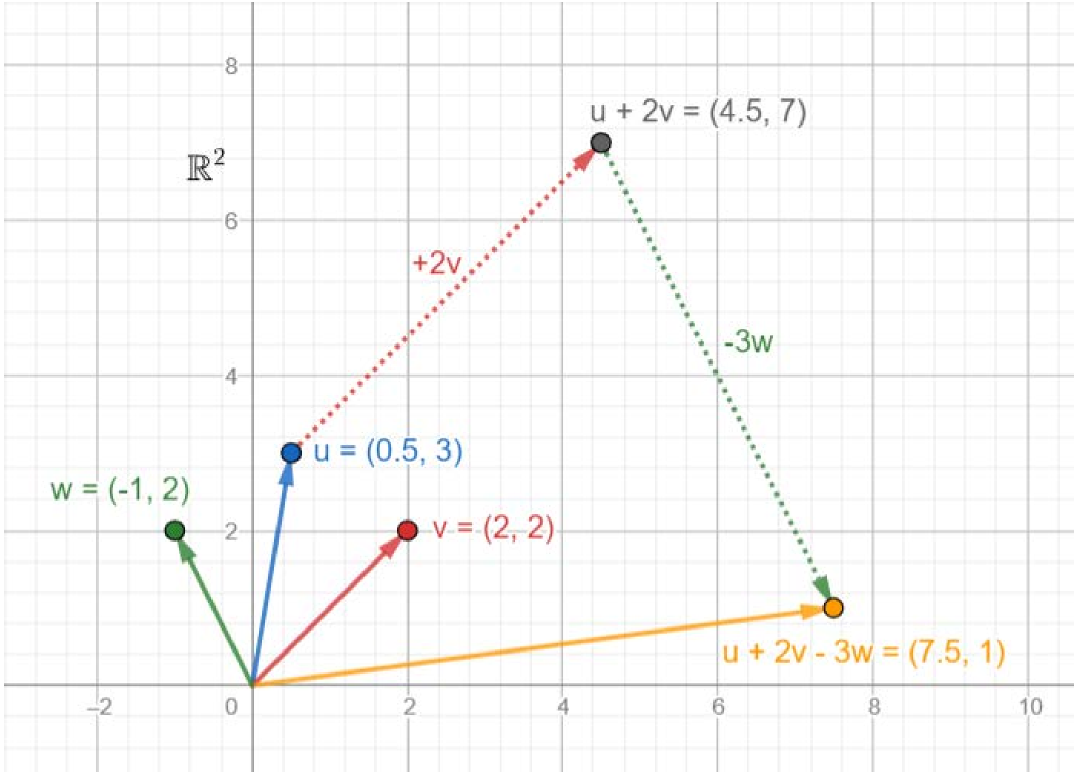
# Linear Combination

- Vectors $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^d$ and scalar coefficients $a_1, \dots, a_k \in \mathbb{R}$

**Linear combination:**

$$\sum_{i=1}^{k} a_i \vec{v}_i = a_1 \vec{v}_1 + \cdots + a_k \vec{v}_k$$

"connecting scaled arrows"



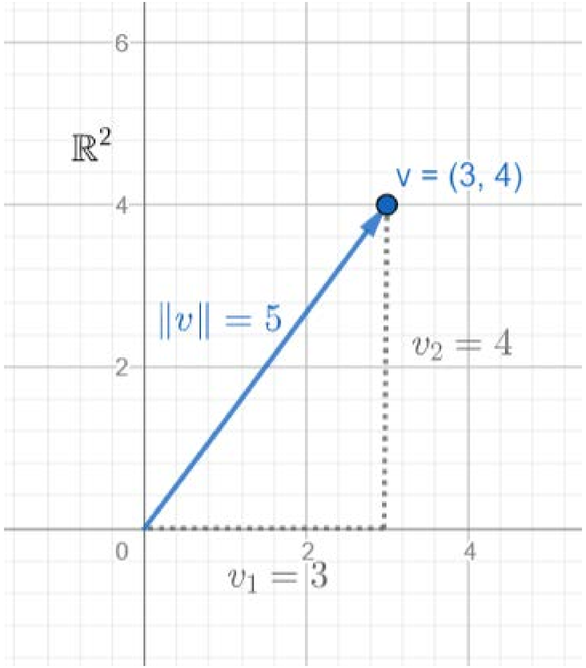
# L2-Norm

- Norm = length of vector = distance from origin
- **Pythagorean theorem:**

$$\vec{v} = (v_1, \dots, v_d) \in \mathbb{R}^d$$

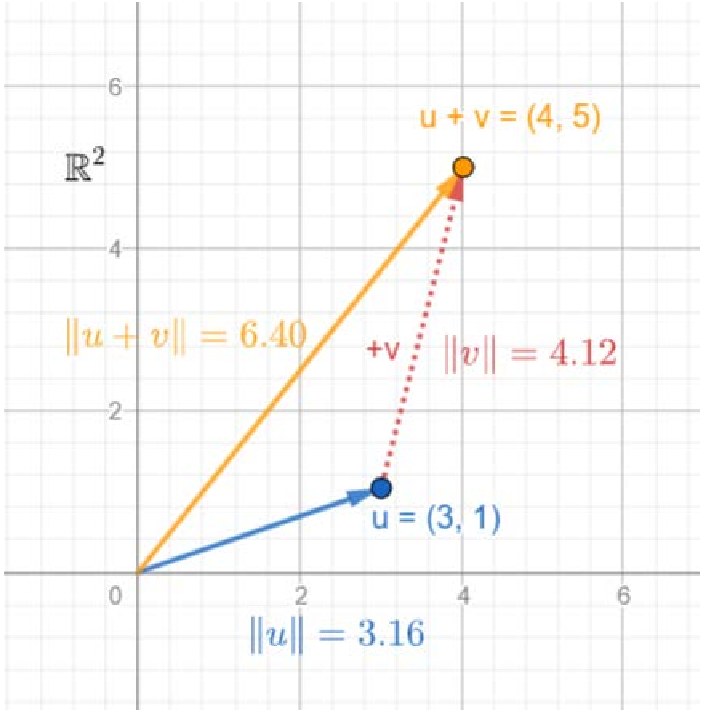
$$\|\vec{v}\| = \sqrt{(v_1)^2 + \cdots + (v_d)^2}$$

- In 1-dim real number line $\mathbb{R}$, norm is just the absolute value $|v|$



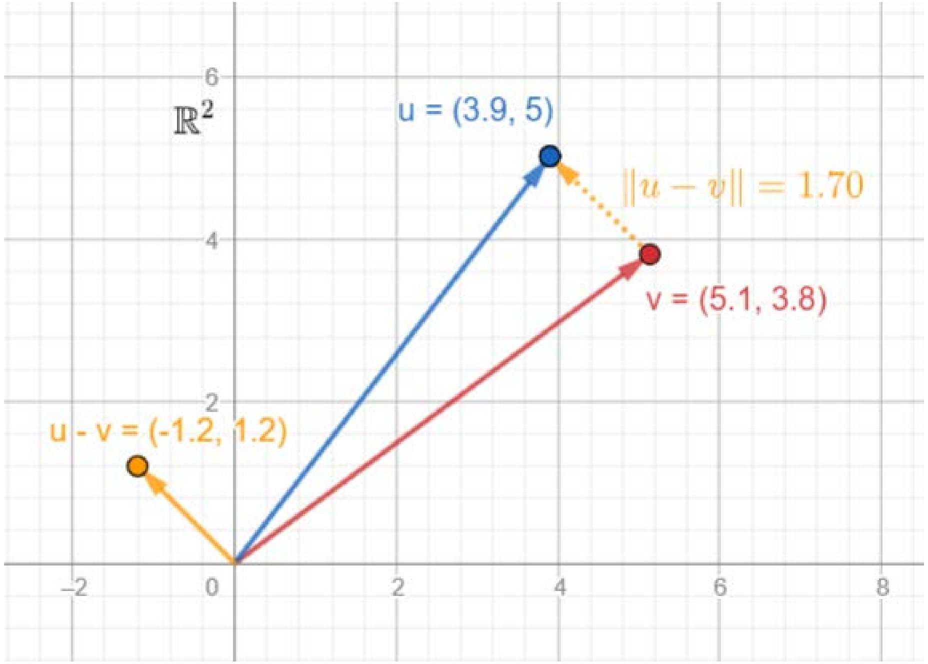
# Norm Properties

**Triangle inequality:** $\|\vec{u}+\vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$




A **measure of similarity** of two vectors:

$\vec{u} \approx \vec{v}$ if $\|\vec{u}-\vec{v}\| \approx 0$



# Unit Vectors and Unit Circle

- **Unit vector:** A vector with norm $\|\vec{v}\| = 1$
- **Standard unit vectors** in $\mathbb{R}^d$: Axis-aligned

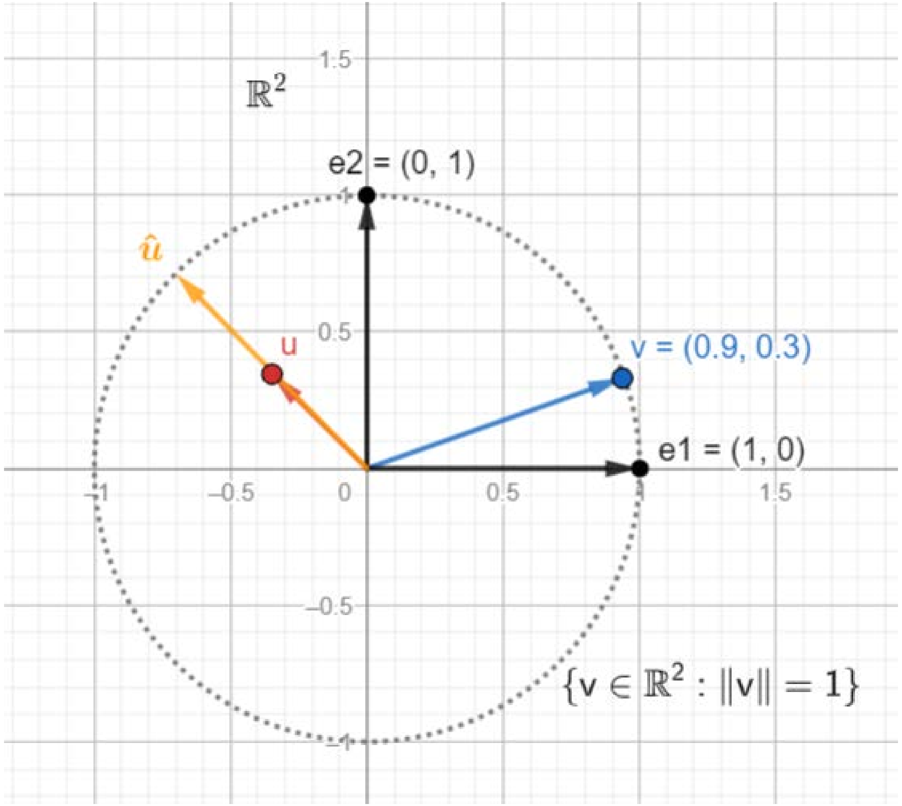
$$\vec{e}_1 = (1,0,\dots,0)$$
$$\dots$$
$$\vec{e}_d = (0,\dots 0,1)$$

- **Normalization:** Can always rescale a non-zero vector to become length 1

$$\hat{v} = \left(\frac{1}{\|\vec{v}\|}\right)\vec{v}$$

- **Unit circle** in $\mathbb{R}^2$ (or unit sphere in $\mathbb{R}^d$):

$$\{\vec{v} \in \mathbb{R}^d : \|\vec{v}\| = 1\}$$



# L1-Norm

- $\boldsymbol{L_1}$**-norm** (also Manhattan / taxi cab norm): Another way to measure vector length

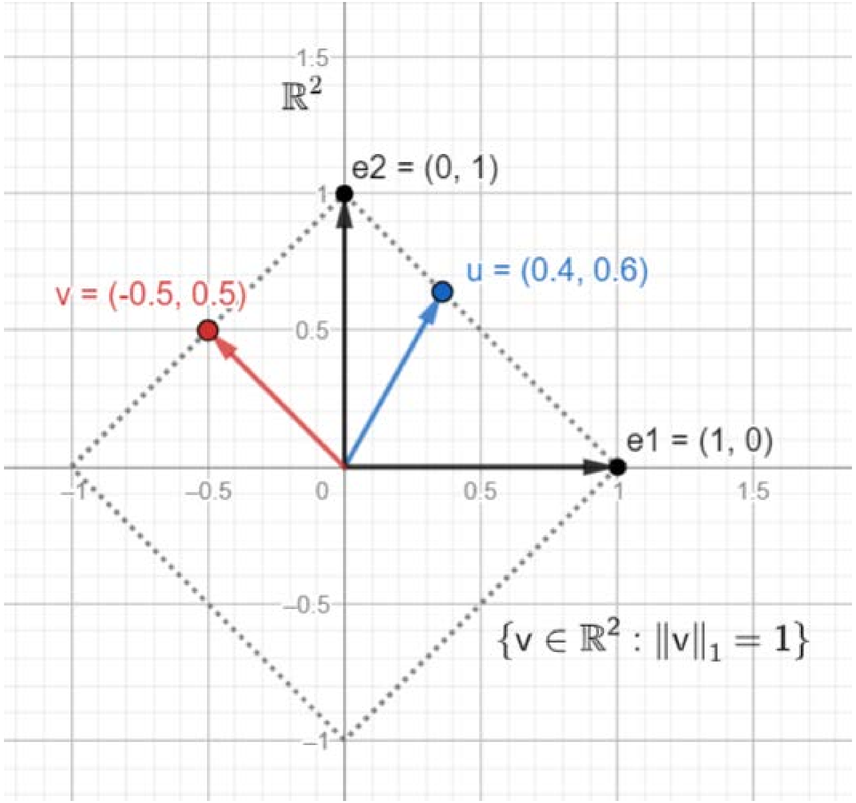
$$\|\vec{v}\|_1 = |v_1| + \cdots + |v_d|$$

(in contrast to previous $L_2$-norm $\|\ldots\|_2$)

- **Unit "circle"**: It has a diamond shape

$$\{\vec{v} \in \mathbb{R}^d : \|\vec{v}\|_1 = 1\}$$

- Application: Regularization of parameters in machine learning models e.g. Lasso regression L1-penalty



# Dot Product

- **Dot product** / scalar product / inner product of 2 vectors:

$\vec{u} = (u_1, \dots, u_d)$ and $\vec{v} = (v_1, \dots, v_d)$

$$\vec{u} \cdot \vec{v} = u_1 \times v_1 + \cdots + u_d \times v_d$$
$$= \sum_{i=1}^{d} u_i v_i$$

e.g. $\vec{u} = (2, 5)$ $\vec{v} = (3, -1)$

$$\vec{u} \cdot \vec{v} = 2 \times 3 + 5 \times (-1) = 1$$

# Dot Product Properties

- Commutative

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

- Distributive

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

- Associative

$$(a\vec{u}) \cdot \vec{v} = a(\vec{u} \cdot \vec{v}) = \vec{u} \cdot (a\vec{v})$$

But $\vec{u} \cdot (\vec{v} \cdot \vec{w})$ is not defined

- Dot product and $L_2$-norm

$$\boxed{\vec{v} \cdot \vec{v} = v_1 v_1 + \cdots + v_d v_d = \|\vec{v}\|^2}$$

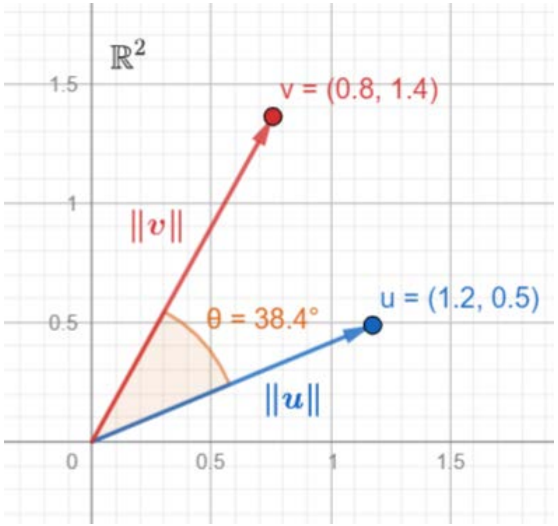
$$\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}$$

# Cosine Angle

- Geometry of dot product: angle $\theta$ between $\vec{u}$ and $\vec{v}$

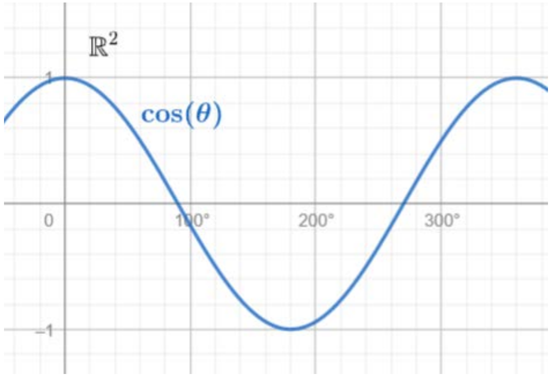
$$\vec{u} \cdot \vec{v} = \underbrace{\|\vec{u}\| \|\vec{v}\|}_{\text{larger if lengthier}} \underbrace{\cos(\theta)}_{\text{always } -1 \le \cdots \le 1}$$

$$\theta = \cos^{-1}\left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|}\right) = \cos^{-1}\left(\frac{\vec{u}}{\|\vec{u}\|}\right) \cdot \left(\frac{\vec{v}}{\|\vec{v}\|}\right)$$




- Unit vectors: If $\|\vec{u}\| = \|\vec{v}\| = 1$, then

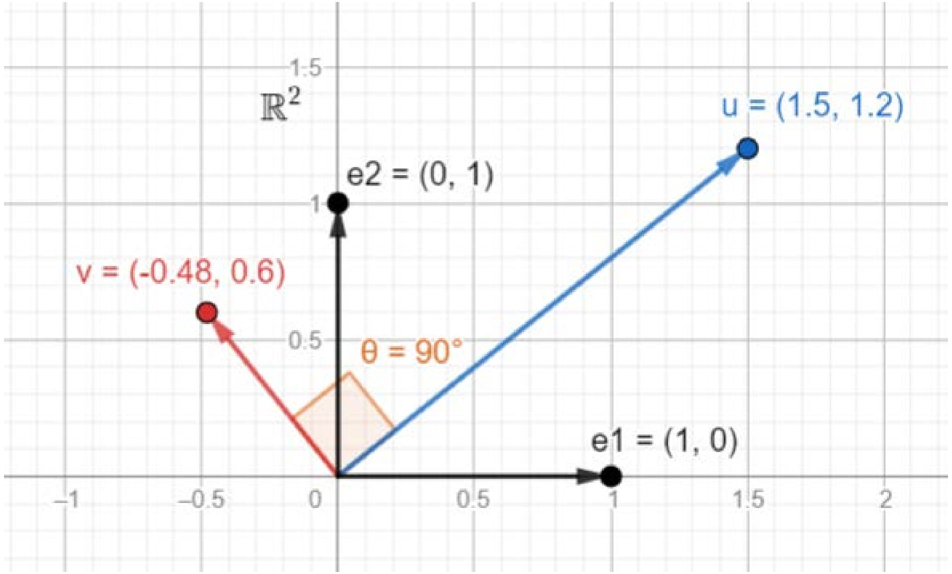
$$\vec{u} \cdot \vec{v} = \cos(\theta)$$



# Orthogonality

- If $\vec{u} \perp \vec{v}$ are perpendicular / orthogonal, then $\theta = 90^\circ$ so $\cos(\theta) = 0$ and

$$\vec{u} \cdot \vec{v} = 0$$




e.g.

$$\vec{u} = (1, 2, 3) \qquad \vec{v} = (10, -5, 0)$$

$$\vec{u} \cdot \vec{v} = 1 \times 10 + 2 \times (-5) + 3 \times 0 = 0$$

# Cosine Similarity

- Vectors $\vec{u}$ and $\vec{v}$ are deemed more similar in the sense that they have **similar direction** even though the norms

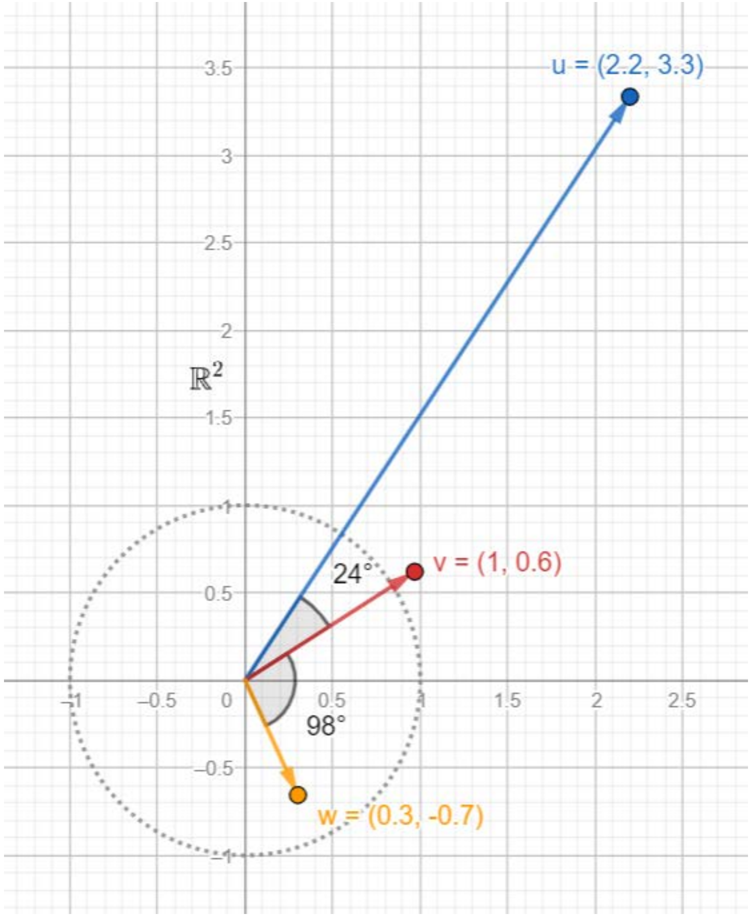
$$||\vec{u} - \vec{v}|| > ||\vec{w} - \vec{v}||$$

- **Angle** $\theta$ between two vectors is another **measure of similarity**. If $\theta$ is small, then

$$\cos(\theta) = \frac{\vec{u} \cdot \vec{v}}{||\vec{u}||\,||\vec{v}||}$$

is large

- Application: Text document vectors similarity in natural language processing NLP



# Linear Independence and Vector Redundancy

- Definition (in English):

  A set $\{\vec{v}_1, \dots, \vec{v}_k\}$ of vectors is **linearly dependent** if some vector can be expressed as a linear combination of the others i.e. there is **redundancy**. Otherwise, the set is **linearly independent**.

e.g. Linearly independent

$$\left\{\begin{pmatrix}-2\\0\\5\end{pmatrix}\right\} \quad \left\{\begin{pmatrix}1\\2\end{pmatrix},\begin{pmatrix}3\\4\end{pmatrix}\right\} \quad \left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\2\\2\end{pmatrix}\right\} \quad \left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\2\\2\end{pmatrix},\begin{pmatrix}1\\2\\3\end{pmatrix}\right\}$$

e.g. Linearly dependent

$$\left\{\begin{pmatrix}1\\-1\\2\end{pmatrix},\begin{pmatrix}-3\\3\\-6\end{pmatrix}\right\} \quad \left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\2\\2\end{pmatrix},\begin{pmatrix}-7\\3\\3\end{pmatrix}\right\} \quad \left\{\begin{pmatrix}1\\2\end{pmatrix},\begin{pmatrix}3\\4\end{pmatrix},\begin{pmatrix}5\\6\end{pmatrix}\right\}$$

# Linear Independence and Non-Trivial Linear Combination

- Definition (in math):

  A set $\{\vec{v}_1, \dots, \vec{v}_k\}$ of vectors is **linearly independent** if the ONLY way to make

  $$a_1\vec{v}_1 + \cdots + a_k\vec{v}_k = \vec{0}$$

  is the obvious trivial way with all $a_1 = \cdots = a_k = 0$. Otherwise, if there is a way to combine the vectors to form $\vec{0}$ with at least one $a_i \neq 0$, then the set is **linearly dependent**.

e.g.

$$\vec{v}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix} \quad \vec{v}_2 = \begin{pmatrix}0\\2\\2\end{pmatrix} \quad \vec{v}_3 = \begin{pmatrix}-7\\3\\3\end{pmatrix}$$

Since $\vec{v}_3 = -7\vec{v}_1 + 1.5\vec{v}_2$, this yields a non-trivial combination $7\vec{v}_1 - 1.5\vec{v}_2 + \vec{v}_3 = \vec{0}$

# Linear Independence Facts

- A **singleton** set $\{\vec{v}\}$ of non-zero vector is always linearly independent
- A set $\{\vec{0}, \vec{v}_2, \dots, \vec{v}_k\}$ with the **zero vector** is always linearly dependent
- A set $\{\vec{v}_1, \vec{v}_2\}$ of **2 vectors** is linearly dependent if they are **parallel** i.e. $\vec{v}_1 = a\vec{v}_2$

$$\vec{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \quad \text{and} \quad \vec{v}_2 = \begin{pmatrix} -3 \\ 3 \\ -6 \end{pmatrix}$$

- If a set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ of **3 vectors** is linearly dependent, then we can write $\vec{v}_3 = a_1\vec{v}_1 + a_2\vec{v}_2$. Geometrically, $\vec{v}_3$ is on the same plane that contains both $\vec{v}_1$ and $\vec{v}_2$ i.e. "**coplanar**"
- **Larger set** of vectors is more likely to have redundancy (linearly dependent)
- In $\mathbb{R}^d$, any set $\{\vec{v}_1, \dots, \vec{v}_{d+1}\}$ of $d + 1$ (or more) vectors is always linearly dependent
- If a smaller set $\{\vec{v}_1, \dots, \vec{v}_k\}$ of vectors is already linearly dependent, then any **extension** to a larger set $\{\vec{v}_1, \dots, \vec{v}_k, \vec{v}_{k+1}, \dots, \vec{v}_{k+n}\}$ must be linearly dependent

## Matrix Definition

- A **matrix** is a rectangular array of numbers

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1c} \\ \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rc} \end{pmatrix} \in \mathbb{R}^{r\times c}$$

$$= \begin{pmatrix} | & & | \\ A_1 & \cdots & A_c \\ | & & | \end{pmatrix}$$

- **Diagonal** elements: $a_{11}, a_{22}, \ldots$

## Row, Column, Square Matrix

- Row vector / matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1c} \end{pmatrix} \in \mathbb{R}^{1\times c}$$

- Column vector / matrix

$$A = \begin{pmatrix} a_{11} \\ \vdots \\ a_{r1} \end{pmatrix} \in \mathbb{R}^{r\times 1}$$

- **Square** matrix: $r = c = d$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{dd} \end{pmatrix} \in \mathbb{R}^{d\times d}$$

## Zero, Diagonal, Identity

- **Zero** matrix $O \in \mathbb{R}^{r \times c}$
- **Diagonal** matrix

$$A = \begin{pmatrix} a_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{dd} \end{pmatrix} \in \mathbb{R}^{d \times d}$$

e.g. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}$

- **Identity** matrix

$$I_d = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{d \times d}$$

## Triangular Matrix

- **Upper triangular** matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{dd} \end{pmatrix} \in \mathbb{R}^{d \times d}$$

e.g. $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{pmatrix}$

- **Lower triangular** matrix

$$A = \begin{pmatrix} a_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{dd} \end{pmatrix} \in \mathbb{R}^{d \times d}$$

## Matrix Addition, Scaling

- Element-wise addition and subtraction

$$A + B = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1c} + b_{1c} \\ \vdots & \ddots & \vdots \\ a_{r1} + b_{r1} & \cdots & a_{rc} + b_{rc} \end{pmatrix}$$

$$A + O = A$$

- Scalar × matrix

$$sA = \begin{pmatrix} s \times a_{11} & \cdots & s \times a_{1c} \\ \vdots & \ddots & \vdots \\ s \times a_{r1} & \cdots & s \times a_{rc} \end{pmatrix}$$

## Matrix Arithmetic Properties

- Commutative

$$A + B = B + A$$

- Associative

$$(A + B) + C = A + (B + C)$$

$$(ab)A = a(bA)$$

- Distributive

$$a(A + B) = aA + aB$$

$$(a + b)A = aA + bA$$

# Matrix Multiplication

- **Shape compatible**?

$$A \in \mathbb{R}^{r\times d} \qquad B \in \mathbb{R}^{d\times c}$$

(number of columns of $A$)

$=$ (number of rows of $B$)

- The **matrix product** is

$$A \times B = AB \in \mathbb{R}^{r\times c}$$

with $ij^{\text{th}}$ element

$$(AB)_{ij} = \sum_{k=1}^{d} A_{ik}B_{kj}$$

$=$ (dot product of $i^{\text{th}}$ row of $A$ and $j^{\text{th}}$ column of $B$)

e.g.

$$A = \begin{pmatrix} 1 & 3 & 0 \\ -2 & 5 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & 0 \\ 7 & 0 & 6 \end{pmatrix}$$

$$AB = \begin{pmatrix} 1 & -7 & 3 \\ 12 & -19 & 6 \end{pmatrix}$$

- Dot product

$$(u_1, \dots, u_d) \times \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix} = \sum_{i=1}^{d} u_i v_i$$

$$= \vec{u} \cdot \vec{v} \in \mathbb{R}^{1\times 1}$$

# Matrix Multiplication Special Cases

- Zero matrix

$$A \in \mathbb{R}^{r\times c} \quad O \in \mathbb{R}^{c\times c} \quad AO \in \mathbb{R}^{r\times c}$$

- Identity matrix

$$A \in \mathbb{R}^{r\times c} \quad I_r A = A \quad AI_c = A$$

- Diagonal matrices

$$\begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_d \end{pmatrix} \times \begin{pmatrix} b_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_d \end{pmatrix} = \begin{pmatrix} a_1 b_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_d b_d \end{pmatrix}$$

- Diagonal matrix

$$\begin{pmatrix} | & & | \\ A_1 & \cdots & A_c \\ | & & | \end{pmatrix} \times \begin{pmatrix} b_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_d \end{pmatrix} = \begin{pmatrix} | & & | \\ b_1 A_1 & \cdots & b_d A_c \\ | & & | \end{pmatrix}$$

e.g.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \times \begin{pmatrix} 10 & 0 \\ 0 & 100 \end{pmatrix} = \begin{pmatrix} 10 & 200 \\ 30 & 400 \end{pmatrix}$$

# Matrix Multiplication Special Cases

- Outer product

$$\begin{pmatrix} u_1 \\ \vdots \\ u_r \end{pmatrix} \times (v_1, \dots, v_c)$$

$$= \begin{pmatrix} u_1 v_1 & \cdots & u_1 v_c \\ \vdots & \ddots & \vdots \\ u_r v_1 & \cdots & u_r v_c \end{pmatrix} \in \mathbb{R}^{r \times c}$$

e.g.

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \times (10 \quad 100) = \begin{pmatrix} 10 & 100 \\ 20 & 200 \\ 30 & 300 \end{pmatrix}$$

- Matrix × vector = another vector

$$\begin{pmatrix} | & & | \\ A_1 & \cdots & A_c \\ | & & | \end{pmatrix} \times \begin{pmatrix} v_1 \\ \vdots \\ v_c \end{pmatrix}$$

$$= \begin{pmatrix} | \\ v_1 A_1 + \cdots + v_c A_c \\ | \end{pmatrix}$$

e.g.

$$\begin{pmatrix} 5 & 6 & -4 \\ 0 & 3 & 8 \end{pmatrix} \times \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$$

$$= 2\begin{pmatrix} 5 \\ 0 \end{pmatrix} - 1\begin{pmatrix} 6 \\ 3 \end{pmatrix} + 0\begin{pmatrix} -4 \\ 8 \end{pmatrix} = \begin{pmatrix} 4 \\ -3 \end{pmatrix}$$

# Matrix Multiplication Properties

- (Mostly) not commutative

$$\boxed{AB \neq BA}$$

e.g.

"permutation matrix"

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

$$AB = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} \neq BA = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}$$

- Associative

$$(AB)C = A(BC)$$

$$s(AB) = (sA)B = A(sB)$$

- Distributive

$$A(B + C) = AB + AC$$

$$(A + B)(C + D) = AC + AD + BC + BD$$

e.g.

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq O \qquad B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \neq O$$

$$AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = O$$

# Matrix Transpose

- **Transpose** mirrors the matrix along its diagonal

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1c} \\ \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rc} \end{pmatrix} \in \mathbb{R}^{r \times c}$$

$$A^T = \begin{pmatrix} a_{11} & \cdots & a_{r1} \\ \vdots & \ddots & \vdots \\ a_{1c} & \cdots & a_{rc} \end{pmatrix} \in \mathbb{R}^{c \times r}$$

e.g.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

# Matrix Transpose Properties

$$(A^T)^T = A$$

$$(A + B)^T = A^T + B^T$$

- Reverse matrix multiplication ordering

$$\boxed{(AB)^T = B^T A^T} \neq A^T B^T$$

More generally

$$(A_1 \cdots A_k)^T = A_k^T \cdots A_1^T$$

# Symmetric Matrix

- A (square) matrix is **symmetric** if

$$A^T = A$$

e.g.

$$A = \begin{pmatrix} 0 & 1 & 4 \\ 1 & 3 & -2 \\ 4 & -2 & 5 \end{pmatrix}$$

# Symmetric Matrix Properties

- If $A$ and $B$ are symmetric, then so is $A + B$
- For any (even non-squared) matrix $A$,

$$\boxed{A^T A \quad \text{and} \quad A A^T}$$

are always symmetric

e.g.

$$A = \begin{pmatrix} -2 & 0 & 5 \\ 1 & 3 & 4 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 5 & 3 & -6 \\ 3 & 9 & 12 \\ -6 & 12 & 41 \end{pmatrix} \qquad A A^T = \begin{pmatrix} 29 & 18 \\ 18 & 26 \end{pmatrix}$$

# Matrix Trace

- **Trace** of a square matrix $A \in \mathbb{R}^{d\times d}$ is the sum of its diagonal elements

$$\mathrm{tr}(A) = \sum_{i=1}^{d} a_{ii}$$

e.g.

$$\mathrm{tr}\begin{pmatrix} 0 & 1 & 4 \\ 6 & 3 & -2 \\ 5 & -7 & 8 \end{pmatrix} = 0 + 3 + 8 = 11$$

# Matrix Trace Properties

$$\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$$

$$\mathrm{tr}(sA) = s \times \mathrm{tr}(A)$$

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

# Matrix Inverse: Existence

- Given a square matrix $A \in \mathbb{R}^{d \times d}$, can we find $B \in \mathbb{R}^{d \times d}$ such that

$$BA = I_d \text{ ?}$$

- Not always possible e.g. when $A = O$

$$B \times O \neq I_d$$

$$? \times 0 \neq 1$$

- We will outline some necessary and sufficient conditions in a later lecture
- In general, use **Gauss-Jordan method** to compute

# Matrix Inverse: Uniqueness

- If we can find some $B \in \mathbb{R}^{d \times d}$ such that

$$BA = I_d$$

then

1. This $B$ is unique. No other matrix $C$ could make $CA = I_d$
2. Notation: Denote $B$ by $A^{-1}$

$$\boxed{A^{-1}A = I_d}$$

3. Say $A$ is **invertible** / has inverse / is **non-singular**
4. Left-inverse is right-inverse too

$$\boxed{AA^{-1} = I_d}$$

## 2x2 Matrix Inverse

- A matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

is invertible if $ad - bc \neq 0$. In this case, the inverse is

$$A^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

e.g.

$$\begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} 3/2 & -1/2 \\ -2 & 1 \end{pmatrix}$$

## Diagonal Matrix Inverse

- A diagonal matrix $A \in \mathbb{R}^{d\times d}$ is invertible if all $a_{ii} \neq 0$

$$A = \begin{pmatrix} a_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{dd} \end{pmatrix}$$

$$A^{-1} = \begin{pmatrix} 1/a_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/a_{dd} \end{pmatrix}$$

e.g.

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -4 \end{pmatrix}^{-1} = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & -1/4 \end{pmatrix}$$

# Matrix Inverse Properties

- If $A$ is invertible, then so is $A^{-1}$. In this case,

$$(A^{-1})^{-1} = A$$

- If $A$ and $B$ are invertible, then so is $AB$

$$\boxed{(AB)^{-1} = B^{-1}A^{-1}}$$

More generally

$$(A_1 \cdots A_k)^{-1} = A_k^{-1} \cdots A_1^{-1}$$

$$(sA)^{-1} = (1/s)A^{-1}$$

- Inverse and transpose

$$(A^T)^{-1} = (A^{-1})^T$$

- However

$$(A + B)^{-1} \neq A^{-1} + B^{-1}$$

e.g.

$$\frac{1}{2+3} \neq \frac{1}{2} + \frac{1}{3}$$

# Matrix Determinant

- **Determinant** of a square matrix $A$ is
  1. A number
  2. Notation: $\det(A)$ or $|A|$
  3. Defined by a complicated recursive formula

- **Fact:** $A$ is invertible iff $\det(A) \neq 0$

# 2x2 Matrix Determinant

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$\det(A) = ad - bc$$

e.g.

$$\det\begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix} = 2 \times 3 - 1 \times 4 = 2 \neq 0$$

Therefore, the matrix $\begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}$ is invertible.

## Triangular Determinant

- For upper or lower triangular (and hence diagonal) matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{dd} \end{pmatrix} \in \mathbb{R}^{d \times d}$$

$$\det(A) = a_{11} \times \cdots \times a_{dd} = \prod_{i=1}^{d} a_{ii}$$

- If some $a_{ii} = 0$, then $\det(A) = 0$ and $A$ would be singular
- Identity matrix

$$\det(I_d) = 1$$

## Matrix Determinant Properties

$$\det(A^T) = \det(A)$$

$$\det(sA) = s^d \det(A)$$

$$\det(AB) = \det(A)\det(B)$$

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

- Proof?

$$\begin{aligned} 1 = \det(I_d) &= \det(A^{-1}A) \\ &= \det(A^{-1})\det(A) \end{aligned}$$

# Matrix Determinant via Cofactor Expansion

- In general, to compute determinant

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \cdots & a_{dd} \end{pmatrix} \in \mathbb{R}^{d \times d}$$

Use recursive formula

$$\det(A) = \sum_{i=1}^{d} (-1)^{i-1} a_{1i} \det(A[-1, -i])$$

signed

first row of $A$

sub-matrix without 1st row and $i^{\text{th}}$ column

e.g.

$$\det\begin{pmatrix} 1 & 0 & 4 \\ -2 & 3 & 8 \\ 6 & -5 & 7 \end{pmatrix}$$

$$= 1 \times \det\begin{pmatrix} 3 & 8 \\ -5 & 7 \end{pmatrix}$$

$$-0 \times \det\begin{pmatrix} -2 & 8 \\ 6 & 7 \end{pmatrix}$$

$$+4 \times \det\begin{pmatrix} -2 & 3 \\ 6 & -5 \end{pmatrix}$$

$$= 1 \times 61 - 0 \times (-62) + 4 \times (-8)$$

$$= 29$$

# Orthogonal Matrix

- A square matrix

$$Q = \begin{pmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_d \\ | & & | \end{pmatrix} \in \mathbb{R}^{d \times d}$$

is **orthogonal** if its $d$ column vectors are all:

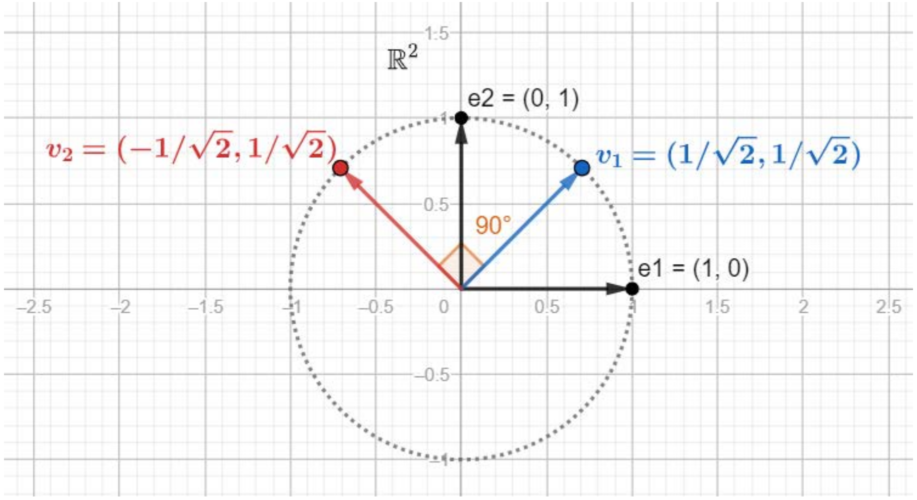
1. Unit vectors $||\vec{v}_i|| = 1$
2. Mutually orthogonal $\vec{v}_i \perp \vec{v}_j$ for $i \neq j$

- Later we see this is same as saying the set $\{\vec{v}_1, \dots, \vec{v}_d\}$ is an **orthonormal basis** for $\mathbb{R}^d$

e.g.

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$



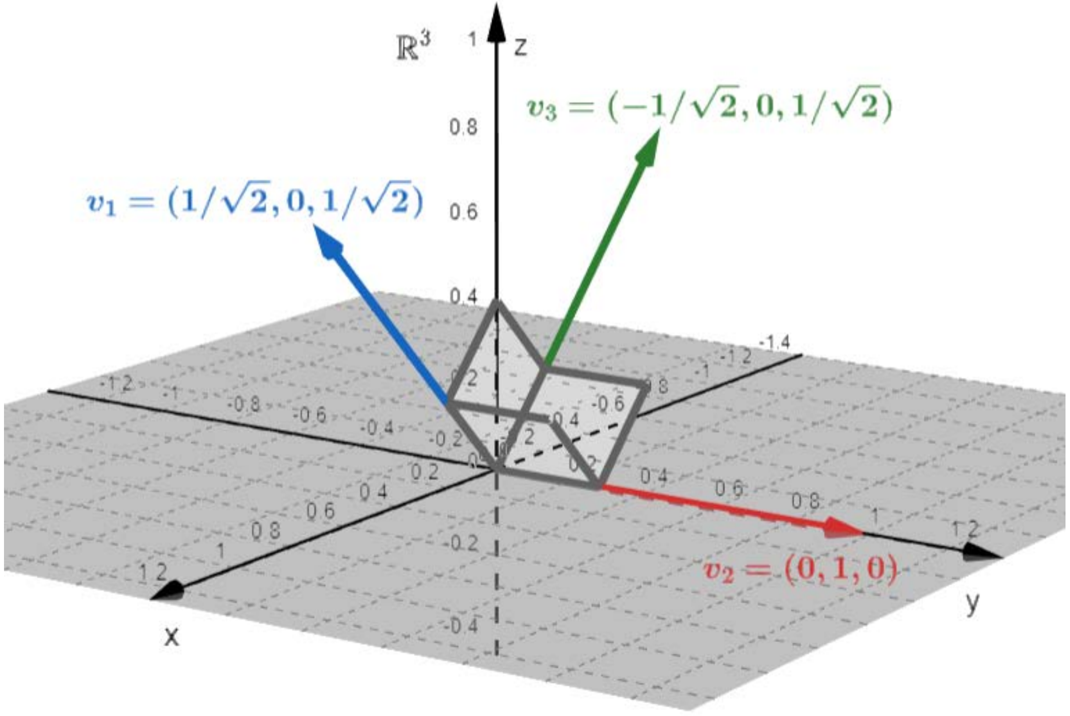
# Orthogonal Matrix

e.g.

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}$$



# Orthogonal Matrix Properties

- If $Q \in \mathbb{R}^{d \times d}$ is an orthogonal matrix, then

$$\vec{v}_i \cdot \vec{v}_i = \|\vec{v}_i\|^2 = 1$$

$$\vec{v}_i \cdot \vec{v}_j = 0 \text{ if } i \neq j$$

$$Q^T Q = \begin{pmatrix} - & \vec{v}_1 & - \\ & \vdots & \\ - & \vec{v}_d & - \end{pmatrix} \begin{pmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_d \\ | & & | \end{pmatrix} = I_d$$

So $Q$ is invertible with a simple inverse

$$\boxed{Q^{-1} = Q^T}$$

- Left-inverse is right-inverse

$$QQ^T = I_d$$

- The $d$ row vectors of $Q$ is also an orthonormal basis for $\mathbb{R}^d$
- Rows of $Q$ are column of $Q^T$. So $Q^T$ is also an orthogonal matrix

- $\det(Q) = +1$ or $-1$

Proof?

$$1 = \det(I_d) = \det(Q^T Q)$$
$$= \det(Q^T)\det(Q) = \det(Q)^2$$