MATH 323

Linear Algebra

Lecture 20:
Diagonalization (continued).
Euclidean structure in $\mathbb{R}^n$.
Orthogonality.

Diagonalization

Theorem 1 Let L be a linear operator on a finite-dimensional vector space V. Then the following conditions are equivalent:

- the matrix of L with respect to some basis is diagonal;
- there exists a basis for V formed by eigenvectors of L.

The operator L is **diagonalizable** if it satisfies these conditions.

Theorem 2 Let A be an $n \times n$ matrix. Then the following conditions are equivalent:

- A is the matrix of a diagonalizable operator;
- A is similar to a diagonal matrix, i.e., it is represented as $A = UBU^{-1}$, where the matrix B is diagonal;
- there exists a basis for $\mathbb{R}^n$ formed by eigenvectors of A.

The matrix A is **diagonalizable** if it satisfies these conditions.

To *diagonalize* an $n \times n$ matrix A is to find a diagonal matrix B and an invertible matrix U such that $A = UBU^{-1}$.

Suppose there exists a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ for $\mathbb{R}^n$ consisting of eigenvectors of A. That is, $A\mathbf{v}_k = \lambda_k \mathbf{v}_k$, where $\lambda_k \in \mathbb{R}$.
Then $A = UBU^{-1}$, where $B = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ and U is a transition matrix whose columns are vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.

Example. $A = \begin{pmatrix} 4 & 3 \\ 0 & 1 \end{pmatrix}$. $\quad \det(A - \lambda I) = (4 - \lambda)(1 - \lambda)$.

Eigenvalues: $\lambda_1 = 4$, $\lambda_2 = 1$.
Associated eigenvectors: $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Thus $A = UBU^{-1}$, where

$$B = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

Suppose we have a problem that involves a square matrix A in the context of matrix multiplication.

Also, suppose that the case when A is a diagonal matrix is simple. Then the diagonalization may help in solving this problem (or may not). Namely, it may reduce the case of a diagonalizable matrix to that of a diagonal one.

An example of such a problem is, given a square matrix A, to find its power A^k:

$$A = \begin{pmatrix} s_1 & & & O \\ & s_2 & & \\ & & \ddots & \\ O & & & s_n \end{pmatrix} \implies A^k = \begin{pmatrix} s_1^k & & & O \\ & s_2^k & & \\ & & \ddots & \\ O & & & s_n^k \end{pmatrix}$$

Problem. Let $A = \begin{pmatrix} 4 & 3 \\ 0 & 1 \end{pmatrix}$. Find A^5.

We know that $A = UBU^{-1}$, where

$$B = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

Then $A^5 = UBU^{-1}UBU^{-1}UBU^{-1}UBU^{-1}UBU^{-1}$

$$= UB^5U^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1024 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1024 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1024 & 1023 \\ 0 & 1 \end{pmatrix}.$$

Problem. Let $A = \begin{pmatrix} 4 & 3 \\ 0 & 1 \end{pmatrix}$. Find A^k ($k \geq 1$).

We know that $A = UBU^{-1}$, where

$$B = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

Then

$$A^k = UB^kU^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 4^k & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 4^k & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 4^k & 4^k - 1 \\ 0 & 1 \end{pmatrix}.$$

Problem. Let $A = \begin{pmatrix} 4 & 3 \\ 0 & 1 \end{pmatrix}$. Find a matrix C such that $C^2 = A$.

We know that $A = UBU^{-1}$, where

$$B = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

Suppose that $D^2 = B$ for some matrix D. Let $C = UDU^{-1}$.
Then $C^2 = UDU^{-1}UDU^{-1} = UD^2U^{-1} = UBU^{-1} = A$.

We can take $D = \begin{pmatrix} \sqrt{4} & 0 \\ 0 & \sqrt{1} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$.

Then $C = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$.

Initial value problem for a system of linear ODEs:

$$\begin{cases} \frac{dx}{dt} = 4x + 3y, \\ \frac{dy}{dt} = y, \end{cases} \qquad x(0) = 1, \ \ y(0) = 1.$$

The system can be rewritten in vector form:

$$\frac{d\mathbf{v}}{dt} = A\mathbf{v}, \quad \text{where} \ \ A = \begin{pmatrix} 4 & 3 \\ 0 & 1 \end{pmatrix}, \ \ \mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Matrix A is diagonalizable: $A = UBU^{-1}$, where

$$B = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}, \qquad U = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

Let $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ be coordinates of the vector $\mathbf{v}$ relative to the basis $\mathbf{v}_1 = (1, 0)$, $\mathbf{v}_2 = (-1, 1)$ of eigenvectors of A. Then $\mathbf{v} = U\mathbf{w} \implies \mathbf{w} = U^{-1}\mathbf{v}$.

It follows that

$$\frac{d\mathbf{w}}{dt} = \frac{d}{dt}(U^{-1}\mathbf{v}) = U^{-1}\frac{d\mathbf{v}}{dt} = U^{-1}A\mathbf{v} = U^{-1}AU\mathbf{w}.$$

Hence $\quad \dfrac{d\mathbf{w}}{dt} = B\mathbf{w} \quad \iff \quad \begin{cases} \frac{dw_1}{dt} = 4w_1, \\ \frac{dw_2}{dt} = w_2. \end{cases}$

General solution: $w_1(t) = c_1 e^{4t}$, $w_2(t) = c_2 e^{t}$, where $c_1, c_2 \in \mathbb{R}$.

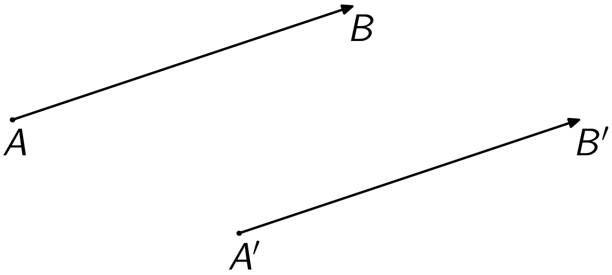
Initial condition:

$$\mathbf{w}(0) = U^{-1}\mathbf{v}(0) = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Thus $w_1(t) = 2e^{4t}$, $w_2(t) = e^{t}$. Then

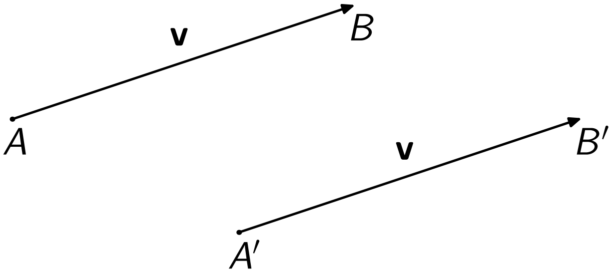
$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = U\mathbf{w}(t) = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2e^{4t} \\ e^{t} \end{pmatrix} = \begin{pmatrix} 2e^{4t} - e^{t} \\ e^{t} \end{pmatrix}.$$

Vectors: geometric approach



- A vector is represented by a directed segment.
- Directed segment is drawn as an arrow.
- Different arrows represent the same vector if they are of the same length and direction.

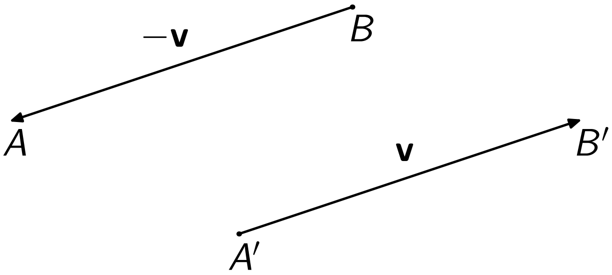
Vectors: geometric approach



$\overrightarrow{AB}$ denotes the vector represented by the arrow with tip at B and tail at A.

$\overrightarrow{AA}$ is called the *zero vector* and denoted $\mathbf{0}$.

Vectors: geometric approach

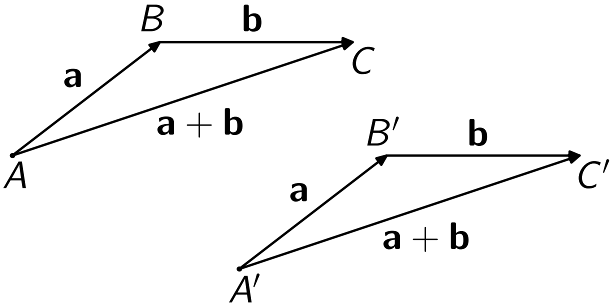


If $\mathbf{v} = \overrightarrow{AB}$ then $\overrightarrow{BA}$ is called the *negative vector* of $\mathbf{v}$ and denoted $-\mathbf{v}$.

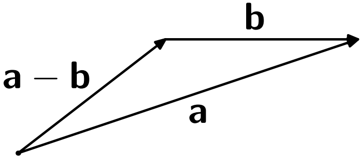
Linear structure: vector addition

Given vectors $\mathbf{a}$ and $\mathbf{b}$, their *sum* $\mathbf{a}+\mathbf{b}$ is defined by the rule $\overrightarrow{AB}+\overrightarrow{BC}=\overrightarrow{AC}$.

That is, choose points A, B, C so that $\overrightarrow{AB}=\mathbf{a}$ and $\overrightarrow{BC}=\mathbf{b}$. Then $\mathbf{a}+\mathbf{b}=\overrightarrow{AC}$.

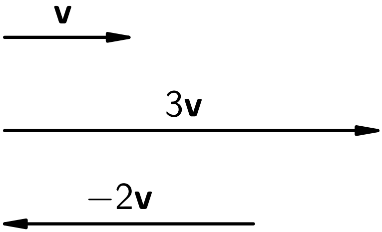


The *difference* of the two vectors is defined as
$\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b})$.



Linear structure: scalar multiplication

Let $\mathbf{v}$ be a vector and $r \in \mathbb{R}$. By definition, $r\mathbf{v}$ is a vector whose magnitude is $|r|$ times the magnitude of $\mathbf{v}$. The direction of $r\mathbf{v}$ coincides with that of $\mathbf{v}$ if $r > 0$. If $r < 0$ then the directions of $r\mathbf{v}$ and $\mathbf{v}$ are opposite.



Beyond linearity: length of a vector

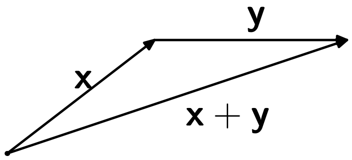
The **length** (or the **magnitude**) of a vector $\overrightarrow{AB}$ is the length of the representing segment AB. The length of a vector $\mathbf{v}$ is denoted $|\mathbf{v}|$ or $\|\mathbf{v}\|$.

Properties of vector length:

$|\mathbf{x}| \geq 0,\ \ |\mathbf{x}| = 0$ only if $\mathbf{x} = \mathbf{0}$ (positivity)

$|r\mathbf{x}| = |r|\,|\mathbf{x}|$ (homogeneity)

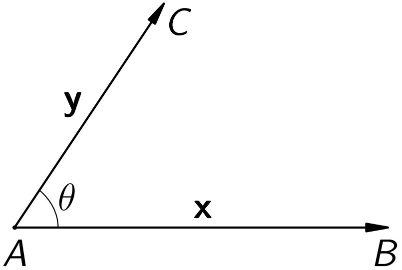
$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$ (triangle inequality)

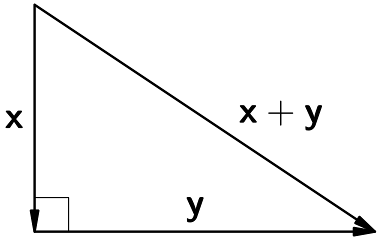


Beyond linearity: angle between vectors

Given nonzero vectors $\mathbf{x}$ and $\mathbf{y}$, let A, B, and C be points such that $\overrightarrow{AB} = \mathbf{x}$ and $\overrightarrow{AC} = \mathbf{y}$. Then $\angle BAC$ is called the **angle** between $\mathbf{x}$ and $\mathbf{y}$.

The vectors $\mathbf{x}$ and $\mathbf{y}$ are called **orthogonal** (denoted $\mathbf{x} \perp \mathbf{y}$) if the angle between them equals 90°.





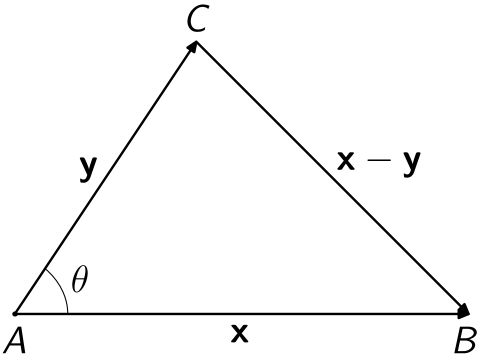
Pythagorean Theorem:

$$\mathbf{x} \perp \mathbf{y} \implies |\mathbf{x}+\mathbf{y}|^2 = |\mathbf{x}|^2 + |\mathbf{y}|^2$$

3-*dimensional Pythagorean Theorem:*

If vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are pairwise orthogonal then

$$|\mathbf{x}+\mathbf{y}+\mathbf{z}|^2 = |\mathbf{x}|^2 + |\mathbf{y}|^2 + |\mathbf{z}|^2$$



Law of cosines:

$$|\mathbf{x} - \mathbf{y}|^2 = |\mathbf{x}|^2 + |\mathbf{y}|^2 - 2|\mathbf{x}|\,|\mathbf{y}|\,\cos\theta$$

Beyond linearity: dot product

The **dot product** of vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}|\,|\mathbf{y}|\,\cos\theta,$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$.

The dot product is also called the **scalar product**.
Alternative notation: $(\mathbf{x}, \mathbf{y})$ or $\langle \mathbf{x}, \mathbf{y} \rangle$.

Nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if and only if $\mathbf{x} \cdot \mathbf{y} = 0$.

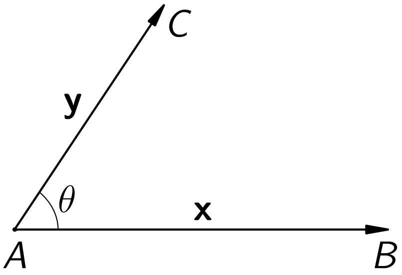
Relations between lengths and dot products:

- $|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$
- $|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}|\,|\mathbf{y}|$
- $|\mathbf{x} - \mathbf{y}|^2 = |\mathbf{x}|^2 + |\mathbf{y}|^2 - 2\,\mathbf{x} \cdot \mathbf{y}$

Euclidean structure

Euclidean structure includes:

- length of a vector: $|\mathbf{x}|$,
- angle between vectors: θ,
- dot product: $\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}|\,|\mathbf{y}|\,\cos\theta$.



Vectors: algebraic approach

An n-dimensional coordinate vector is an element of $\mathbb{R}^n$, i.e., an ordered n-tuple $(x_1, x_2, \dots, x_n)$ of real numbers.

Let $\mathbf{a} = (a_1, a_2, \dots, a_n)$ and $\mathbf{b} = (b_1, b_2, \dots, b_n)$ be vectors, and $r \in \mathbb{R}$ be a scalar. Then, by definition,

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n),$$

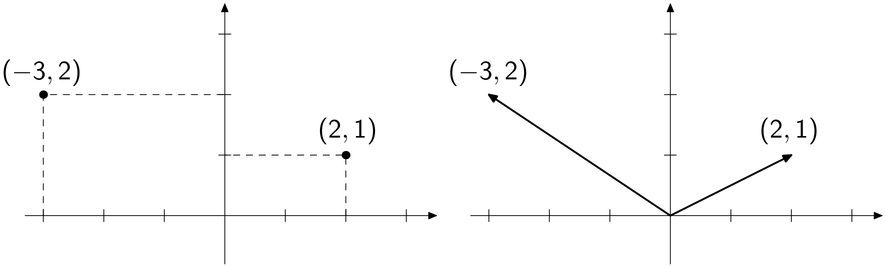
$$r\mathbf{a} = (ra_1, ra_2, \dots, ra_n),$$

$$\mathbf{0} = (0, 0, \dots, 0),$$

$$-\mathbf{b} = (-b_1, -b_2, \dots, -b_n),$$

$$\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b}) = (a_1 - b_1, a_2 - b_2, \dots, a_n - b_n).$$

Cartesian coordinates: geometric meets algebraic



Cartesian coordinates allow us to identify a line, a plane, and space with $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$, respectively.

Once we specify an *origin* O, each point A is associated a *position vector* $\overrightarrow{OA}$. Conversely, every vector has a unique representative with tail at O.

Length and distance

Definition. The **length** of a vector $\mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$ is

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

The **distance** between vectors **x** and **y** is defined as $\|\mathbf{y} - \mathbf{x}\|$.

Properties of length:

$\|\mathbf{x}\| \geq 0$, $\|\mathbf{x}\| = 0$ only if $\mathbf{x} = \mathbf{0}$ (positivity)

$\|r\mathbf{x}\| = |r|\, \|\mathbf{x}\|$ (homogeneity)

$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (triangle inequality)

Scalar product

Definition. The **scalar product** of vectors $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$ is

$$\boxed{\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.}$$

Alternative notation: $(\mathbf{x}, \mathbf{y})$ or $\langle \mathbf{x}, \mathbf{y} \rangle$.

Properties of scalar product:

$\mathbf{x} \cdot \mathbf{x} \geq 0$, $\mathbf{x} \cdot \mathbf{x} = 0$ only if $\mathbf{x} = \mathbf{0}$ (positivity)

$\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ (symmetry)

$(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$ (distributive law)

$(r\mathbf{x}) \cdot \mathbf{y} = r(\mathbf{x} \cdot \mathbf{y})$ (homogeneity)

In particular, $\mathbf{x} \cdot \mathbf{y}$ is a **bilinear** function (i.e., it is both a linear function of $\mathbf{x}$ and a linear function of $\mathbf{y}$).

Angle

Cauchy-Schwarz inequality: $|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|.$

By the Cauchy-Schwarz inequality, for any nonzero vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we have

$$\cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \, \|\mathbf{y}\|} \quad \text{for a unique} \quad 0 \leq \theta \leq \pi.$$

θ is called the **angle** between the vectors **x** and **y**.

The vectors **x** and **y** are said to be **orthogonal** (denoted $\mathbf{x} \perp \mathbf{y}$) if $\mathbf{x} \cdot \mathbf{y} = 0$ (i.e., if $\theta = 90^{\circ}$).

Problem. Find the angle θ between vectors $\mathbf{x} = (2, -1)$ and $\mathbf{y} = (3, 1)$.

$\mathbf{x} \cdot \mathbf{y} = 5$, $\|\mathbf{x}\| = \sqrt{5}$, $\|\mathbf{y}\| = \sqrt{10}$.

$$\cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \, \|\mathbf{y}\|} = \frac{5}{\sqrt{5}\,\sqrt{10}} = \frac{1}{\sqrt{2}} \implies \theta = 45^{\text{o}}$$

Problem. Find the angle ϕ between vectors $\mathbf{v} = (-2, 1, 3)$ and $\mathbf{w} = (4, 5, 1)$.

$$\mathbf{v} \cdot \mathbf{w} = 0 \implies \mathbf{v} \perp \mathbf{w} \implies \phi = 90^{\text{o}}$$

Orthogonality

Definition 1. Vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are said to be **orthogonal** (denoted $\mathbf{x} \perp \mathbf{y}$) if $\boxed{\mathbf{x} \cdot \mathbf{y} = 0.}$

Definition 2. A vector $\mathbf{x} \in \mathbb{R}^n$ is said to be **orthogonal** to a nonempty set $Y \subset \mathbb{R}^n$ (denoted $\mathbf{x} \perp Y$) if $\mathbf{x} \cdot \mathbf{y} = 0$ for any $\mathbf{y} \in Y$.

Definition 3. Nonempty sets $X, Y \subset \mathbb{R}^n$ are said to be **orthogonal** (denoted $X \perp Y$) if $\mathbf{x} \cdot \mathbf{y} = 0$ for any $\mathbf{x} \in X$ and $\mathbf{y} \in Y$.

Examples in $\mathbb{R}^3$. • The line $x = y = 0$ is orthogonal to the line $y = z = 0$.

Indeed, if $\mathbf{v} = (0, 0, z)$ and $\mathbf{w} = (x, 0, 0)$ then $\mathbf{v} \cdot \mathbf{w} = 0$.

• The line $x = y = 0$ is orthogonal to the plane $z = 0$.

Indeed, if $\mathbf{v} = (0, 0, z)$ and $\mathbf{w} = (x, y, 0)$ then $\mathbf{v} \cdot \mathbf{w} = 0$.

• The line $x = y = 0$ is not orthogonal to the plane $z = 1$.

The vector $\mathbf{v} = (0, 0, 1)$ belongs to both the line and the plane, and $\mathbf{v} \cdot \mathbf{v} = 1 \neq 0$.

• The plane $z = 0$ is not orthogonal to the plane $y = 0$.

The vector $\mathbf{v} = (1, 0, 0)$ belongs to both planes and $\mathbf{v} \cdot \mathbf{v} = 1 \neq 0$.

Proposition 1 If $X, Y \in \mathbb{R}^n$ are orthogonal sets then either they are disjoint or $X \cap Y = \{\mathbf{0}\}$.

Proof: $\mathbf{v} \in X \cap Y \implies \mathbf{v} \perp \mathbf{v} \implies \mathbf{v} \cdot \mathbf{v} = 0 \implies \mathbf{v} = \mathbf{0}$.

Proposition 2 Let V be a subspace of $\mathbb{R}^n$ and S be a spanning set for V. Then for any $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x} \perp S \implies \mathbf{x} \perp V.$$

Proof: Any $\mathbf{v} \in V$ is represented as $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k$, where $\mathbf{v}_i \in S$ and $a_i \in \mathbb{R}$. If $\mathbf{x} \perp S$ then

$$\mathbf{x} \cdot \mathbf{v} = a_1(\mathbf{x} \cdot \mathbf{v}_1) + \cdots + a_k(\mathbf{x} \cdot \mathbf{v}_k) = 0 \implies \mathbf{x} \perp \mathbf{v}.$$

Example. The vector $\mathbf{v} = (1, 1, 1)$ is orthogonal to the plane spanned by vectors $\mathbf{w}_1 = (2, -3, 1)$ and $\mathbf{w}_2 = (0, 1, -1)$ (because $\mathbf{v} \cdot \mathbf{w}_1 = \mathbf{v} \cdot \mathbf{w}_2 = 0$).

Orthogonal complement

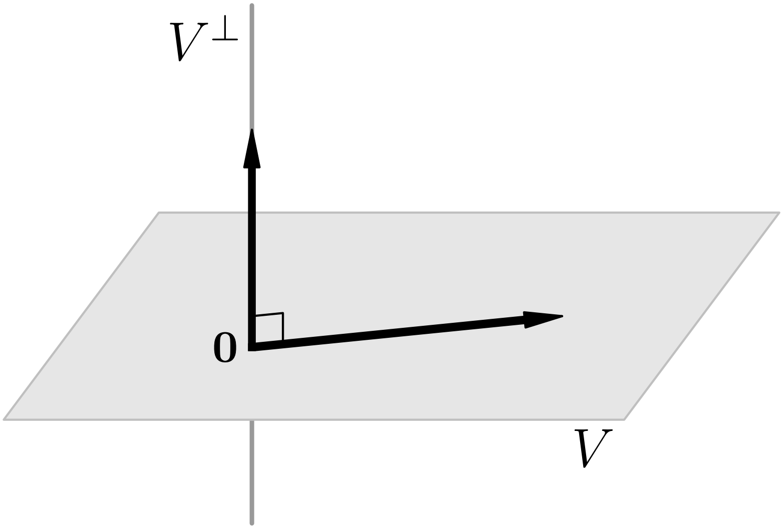
Definition. Let $S \subset \mathbb{R}^n$. The **orthogonal complement** of S, denoted $S^\perp$, is the set of all vectors $\mathbf{x} \in \mathbb{R}^n$ that are orthogonal to S. That is, $S^\perp$ is the largest subset of $\mathbb{R}^n$ orthogonal to S.

Theorem 1 $S^\perp$ is a subspace of $\mathbb{R}^n$.

Note that $S \subset (S^\perp)^\perp$, hence $\mathrm{Span}(S) \subset (S^\perp)^\perp$.

Theorem 2 $(S^\perp)^\perp = \mathrm{Span}(S)$. In particular, for any subspace V we have $(V^\perp)^\perp = V$.

Example. Consider a line $L = \{(x, 0, 0) \mid x \in \mathbb{R}\}$ and a plane $\Pi = \{(0, y, z) \mid y, z \in \mathbb{R}\}$ in $\mathbb{R}^3$. Then $L^\perp = \Pi$ and $\Pi^\perp = L$.



$V^\perp$
$\mathbf{0}$
V