MATH 304

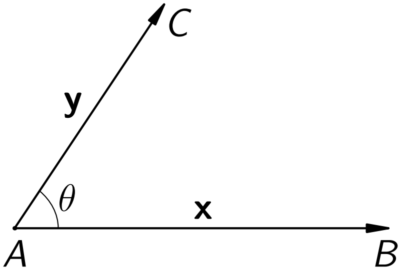
Linear Algebra

Lecture 18:
Orthogonal complement (continued).
Orthogonal projection.
Least squares problems.

Euclidean structure

Euclidean structure in $\mathbb{R}^n$ includes:

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



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/points $\mathbf{x}$ and $\mathbf{y}$ is $\|\mathbf{y} - \mathbf{x}\|$.

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

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

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 $\mathbf{x}$ and $\mathbf{y}$.

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

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 $\mathbf{y} \in Y$.

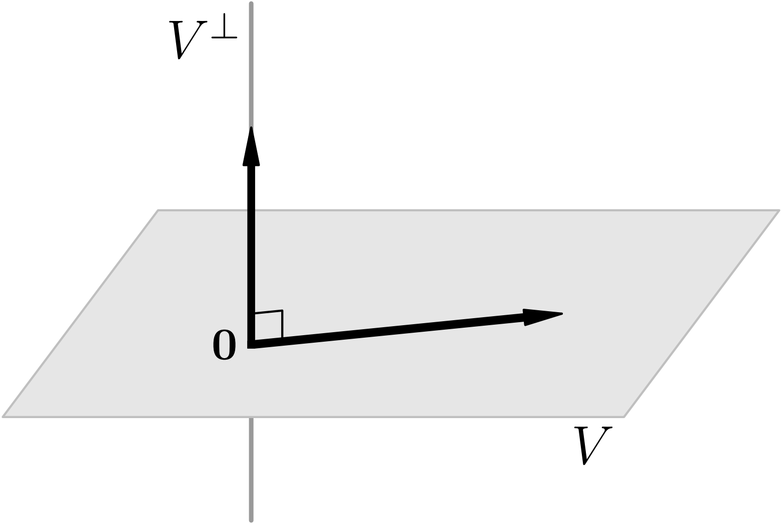
Definition 3. Nonempty sets $X, Y \subset \mathbb{R}^n$ are said to be **orthogonal** (denoted $X \perp Y$) if $\mathbf{x} \cdot \mathbf{y} = 0$ for any $\mathbf{x} \in X$ and $\mathbf{y} \in Y$.

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.

Theorem 1 (i) $S^\perp$ is a subspace of $\mathbb{R}^n$.
(ii) $\mathrm{Span}(S)^\perp = S^\perp$.
(iii) $(S^\perp)^\perp = \mathrm{Span}(S)$.

Theorem 2 If V is a subspace of $\mathbb{R}^n$, then
(i) $(V^\perp)^\perp = V$,
(ii) $V \cap V^\perp = \{\mathbf{0}\}$.



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

Fundamental subspaces

Definition. Given an $m \times n$ matrix A, let

$$N(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\},$$

$$R(A) = \{\mathbf{b} \in \mathbb{R}^m \mid \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^n\}.$$

$R(A)$ is the range of a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$, $L(\mathbf{x}) = A\mathbf{x}$. $N(A)$ is the kernel of L.

Also, $N(A)$ is the nullspace of the matrix A while $R(A)$ is the column space of A. The row space of A is $R(A^T)$.

The subspaces $N(A), R(A^T) \subset \mathbb{R}^n$ and $R(A), N(A^T) \subset \mathbb{R}^m$ are **fundamental subspaces** associated to the matrix A.

Theorem $N(A) = R(A^T)^\perp$, $N(A^T) = R(A)^\perp$.
That is, the nullspace of a matrix is the orthogonal complement of its row space.

Proof: The equality $A\mathbf{x} = \mathbf{0}$ means that the vector $\mathbf{x}$ is orthogonal to rows of the matrix A. Therefore $N(A) = S^\perp$, where S is the set of rows of A. It remains to note that $S^\perp = \mathrm{Span}(S)^\perp = R(A^T)^\perp$.

Corollary Let V be a subspace of $\mathbb{R}^n$. Then $\dim V + \dim V^\perp = n$.

Proof: Pick a basis $\mathbf{v}_1, \dots, \mathbf{v}_k$ for V. Let A be the $k \times n$ matrix whose rows are vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. Then $V = R(A^T)$, hence $V^\perp = N(A)$. Consequently, $\dim V$ and $\dim V^\perp$ are rank and nullity of A. Therefore $\dim V + \dim V^\perp$ equals the number of columns of A, which is n.

Orthogonal projection

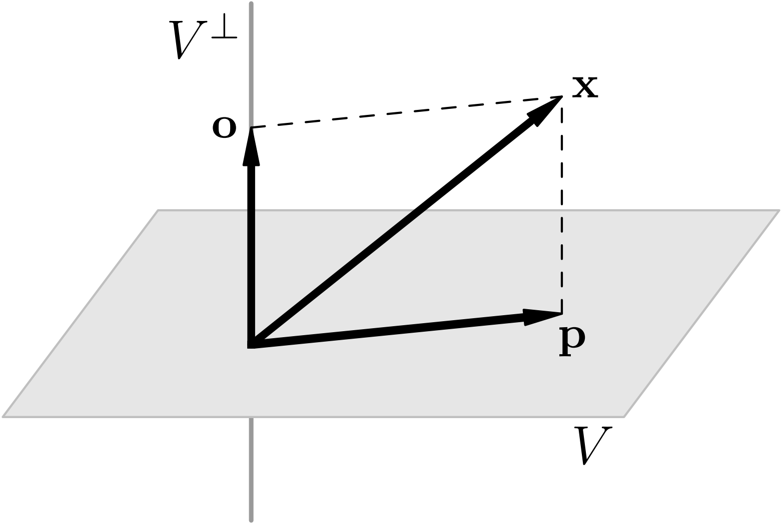
Theorem 1 Let V be a subspace of $\mathbb{R}^n$. Then any vector $\mathbf{x} \in \mathbb{R}^n$ is uniquely represented as $\mathbf{x} = \mathbf{p} + \mathbf{o}$, where $\mathbf{p} \in V$ and $\mathbf{o} \in V^\perp$.

Idea of the proof: Let $\mathbf{v}_1, \dots, \mathbf{v}_k$ be a basis for V and $\mathbf{w}_1, \dots, \mathbf{w}_m$ be a basis for $V^\perp$. Then $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_1, \dots, \mathbf{w}_m$ is a basis for $\mathbb{R}^n$.

In the above expansion, $\mathbf{p}$ is called the **orthogonal projection** of the vector $\mathbf{x}$ onto the subspace V.

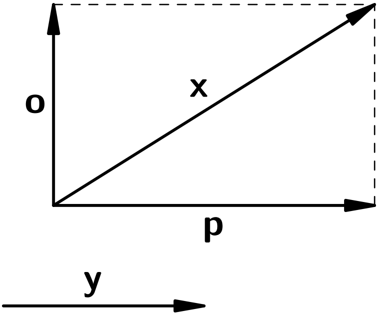
Theorem 2 $\|\mathbf{x} - \mathbf{v}\| > \|\mathbf{x} - \mathbf{p}\|$ for any $\mathbf{v} \neq \mathbf{p}$ in V.

Thus $\|\mathbf{o}\| = \|\mathbf{x} - \mathbf{p}\| = \min_{\mathbf{v} \in V} \|\mathbf{x} - \mathbf{v}\|$ is the **distance** from the vector $\mathbf{x}$ to the subspace V.


$V^\perp$
$\mathbf{x}$
$\mathbf{o}$
$\mathbf{p}$
V

Orthogonal projection onto a vector

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, with $\mathbf{y} \neq \mathbf{0}$.
Then there exists a unique decomposition $\mathbf{x} = \mathbf{p} + \mathbf{o}$ such that $\mathbf{p}$ is parallel to $\mathbf{y}$ and $\mathbf{o}$ is orthogonal to $\mathbf{y}$.



$\mathbf{p}$ = orthogonal projection of $\mathbf{x}$ onto $\mathbf{y}$

Orthogonal projection onto a vector

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, with $\mathbf{y} \neq \mathbf{0}$.
Then there exists a unique decomposition $\mathbf{x} = \mathbf{p} + \mathbf{o}$ such that $\mathbf{p}$ is parallel to $\mathbf{y}$ and $\mathbf{o}$ is orthogonal to $\mathbf{y}$.

We have $\mathbf{p} = \alpha\mathbf{y}$ for some $\alpha \in \mathbb{R}$. Then

$$0 = \mathbf{o} \cdot \mathbf{y} = (\mathbf{x} - \alpha\mathbf{y}) \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{y} - \alpha\mathbf{y} \cdot \mathbf{y}.$$

$$\implies \quad \alpha = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{y} \cdot \mathbf{y}} \quad \implies \quad \boxed{\mathbf{p} = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{y} \cdot \mathbf{y}}\,\mathbf{y}}$$

Problem. Find the distance from the point $\mathbf{x} = (3, 1)$ to the line spanned by $\mathbf{y} = (2, -1)$.

Consider the decomposition $\mathbf{x} = \mathbf{p} + \mathbf{o}$, where $\mathbf{p}$ is parallel to $\mathbf{y}$ while $\mathbf{o} \perp \mathbf{y}$. The required distance is the length of the orthogonal component $\mathbf{o}$.

$$\mathbf{p} = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{y} \cdot \mathbf{y}}\, \mathbf{y} = \frac{5}{5}\,(2, -1) = (2, -1),$$

$$\mathbf{o} = \mathbf{x} - \mathbf{p} = (3, 1) - (2, -1) = (1, 2), \quad \|\mathbf{o}\| = \sqrt{5}.$$

Problem. Find the point on the line $y = -x$ that is closest to the point $(3, 4)$.

The required point is the projection $\mathbf{p}$ of $\mathbf{v} = (3, 4)$ on the vector $\mathbf{w} = (1, -1)$ spanning the line $y = -x$.

$$\mathbf{p} = \frac{\mathbf{v} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}\, \mathbf{w} = \frac{-1}{2}\,(1, -1) = \left(-\frac{1}{2}, \frac{1}{2}\right)$$

Problem. Let Π be the plane spanned by vectors $\mathbf{v}_1 = (1,1,0)$ and $\mathbf{v}_2 = (0,1,1)$.

(i) Find the orthogonal projection of the vector $\mathbf{x} = (4,0,-1)$ onto the plane Π.

(ii) Find the distance from $\mathbf{x}$ to Π.

We have $\mathbf{x} = \mathbf{p} + \mathbf{o}$, where $\mathbf{p} \in \Pi$ and $\mathbf{o} \perp \Pi$. Then the orthogonal projection of $\mathbf{x}$ onto Π is $\mathbf{p}$ and the distance from $\mathbf{x}$ to Π is $\|\mathbf{o}\|$.

We have $\mathbf{p} = \alpha\mathbf{v}_1 + \beta\mathbf{v}_2$ for some $\alpha, \beta \in \mathbb{R}$. Then $\mathbf{o} = \mathbf{x} - \mathbf{p} = \mathbf{x} - \alpha\mathbf{v}_1 - \beta\mathbf{v}_2$.

$$\begin{cases} \mathbf{o} \cdot \mathbf{v}_1 = 0 \\ \mathbf{o} \cdot \mathbf{v}_2 = 0 \end{cases} \iff \begin{cases} \alpha(\mathbf{v}_1 \cdot \mathbf{v}_1) + \beta(\mathbf{v}_2 \cdot \mathbf{v}_1) = \mathbf{x} \cdot \mathbf{v}_1 \\ \alpha(\mathbf{v}_1 \cdot \mathbf{v}_2) + \beta(\mathbf{v}_2 \cdot \mathbf{v}_2) = \mathbf{x} \cdot \mathbf{v}_2 \end{cases}$$

$$\mathbf{x} = (4,0,-1),\ \ \mathbf{v}_1 = (1,1,0),\ \ \mathbf{v}_2 = (0,1,1)$$

$$\begin{cases} \alpha(\mathbf{v}_1 \cdot \mathbf{v}_1) + \beta(\mathbf{v}_2 \cdot \mathbf{v}_1) = \mathbf{x} \cdot \mathbf{v}_1 \\ \alpha(\mathbf{v}_1 \cdot \mathbf{v}_2) + \beta(\mathbf{v}_2 \cdot \mathbf{v}_2) = \mathbf{x} \cdot \mathbf{v}_2 \end{cases}$$

$$\iff \begin{cases} 2\alpha + \beta = 4 \\ \alpha + 2\beta = -1 \end{cases} \iff \begin{cases} \alpha = 3 \\ \beta = -2 \end{cases}$$

$$\mathbf{p} = 3\mathbf{v}_1 - 2\mathbf{v}_2 = (3,1,-2)$$

$$\mathbf{o} = \mathbf{x} - \mathbf{p} = (1,-1,1)$$

$$\|\mathbf{o}\| = \sqrt{3}$$

Problem. Let Π be the plane spanned by vectors $\mathbf{v}_1 = (1,1,0)$ and $\mathbf{v}_2 = (0,1,1)$.

(i) Find the orthogonal projection of the vector $\mathbf{x} = (4,0,-1)$ onto the plane Π.

(ii) Find the distance from $\mathbf{x}$ to Π.

Alternative solution: We have $\mathbf{x} = \mathbf{p} + \mathbf{o}$, where $\mathbf{p} \in \Pi$ and $\mathbf{o} \perp \Pi$. Then the orthogonal projection of $\mathbf{x}$ onto Π is $\mathbf{p}$ and the distance from $\mathbf{x}$ to Π is $\|\mathbf{o}\|$.

Notice that $\mathbf{o}$ is the orthogonal projection of $\mathbf{x}$ onto the orthogonal complement $\Pi^\perp$. In the previous lecture, we found that $\Pi^\perp$ is the line spanned by the vector $\mathbf{y} = (1,-1,1)$. It follows that

$$\mathbf{o} = \frac{\mathbf{x}\cdot\mathbf{y}}{\mathbf{y}\cdot\mathbf{y}}\,\mathbf{y} = \frac{3}{3}\,(1,-1,1) = (1,-1,1).$$

Then $\mathbf{p} = \mathbf{x} - \mathbf{o} = (4,0,-1) - (1,-1,1) = (3,1,-2)$ and $\|\mathbf{o}\| = \sqrt{3}$.

Overdetermined system of linear equations:

$$\begin{cases} x + 2y = 3 \\ 3x + 2y = 5 \\ x + y = 2.09 \end{cases} \iff \begin{cases} x + 2y = 3 \\ -4y = -4 \\ -y = -0.91 \end{cases}$$

No solution: inconsistent system

Assume that a solution (x_0, y_0) does exist but the system is not quite accurate, namely, there may be some errors in the right-hand sides.

Problem. Find a good approximation of (x_0, y_0).

One approach is the **least squares fit**. Namely, we look for a pair (x, y) that minimizes the sum $(x + 2y - 3)^2 + (3x + 2y - 5)^2 + (x + y - 2.09)^2$.

Least squares solution

System of linear equations:

$$\left\{\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{array}\right. \iff A\mathbf{x} = \mathbf{b}$$

For any $\mathbf{x} \in \mathbb{R}^n$ define a **residual** $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$.

The **least squares solution x** to the system is the one that minimizes $\|r(\mathbf{x})\|$ (or, equivalently, $\|r(\mathbf{x})\|^2$).

$$\|r(\mathbf{x})\|^2 = \sum_{i=1}^{m} (a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - b_i)^2$$

Let A be an $m \times n$ matrix and let $\mathbf{b} \in \mathbb{R}^m$.

Theorem A vector $\hat{\mathbf{x}}$ is a least squares solution of the system $A\mathbf{x} = \mathbf{b}$ if and only if it is a solution of the associated **normal system** $\boxed{A^T A\mathbf{x} = A^T\mathbf{b}.}$

Proof: $A\mathbf{x}$ is an arbitrary vector in $R(A)$, the column space of A. Hence the length of $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$ is minimal if $A\mathbf{x}$ is the orthogonal projection of $\mathbf{b}$ onto $R(A)$. That is, if $r(\mathbf{x})$ is orthogonal to $R(A)$.

We know that $R(A)^\perp = N(A^T)$, the nullspace of the transpose matrix. Thus $\hat{\mathbf{x}}$ is a least squares solution if and only if

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

Problem. Find the least squares solution to

$$\begin{cases} x + 2y = 3 \\ 3x + 2y = 5 \\ x + y = 2.09 \end{cases}$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 2.09 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 5 \\ 2.09 \end{pmatrix}$$

$$\begin{pmatrix} 11 & 9 \\ 9 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 20.09 \\ 18.09 \end{pmatrix} \iff \begin{cases} x = 1 \\ y = 1.01 \end{cases}$$