

MATH 304  
Linear Algebra

**Lecture 18:**  
**Orthogonal projection (continued).**  
**Least squares problems.**  
**Normed vector spaces.**

## Orthogonality

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

*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)**  $(S^\perp)^\perp = \text{Span}(S)$ .

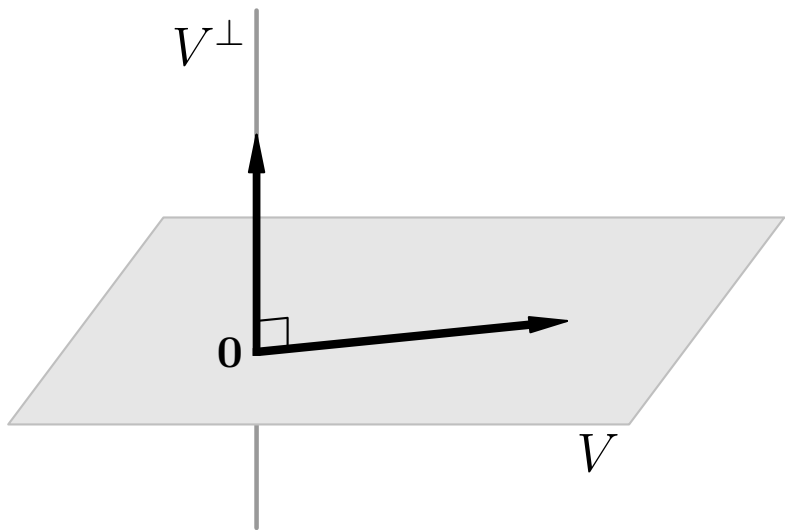
**Theorem 2** If  $V$  is a subspace of  $\mathbb{R}^n$ , then

**(i)**  $(V^\perp)^\perp = V$ ,

**(ii)**  $V \cap V^\perp = \{\mathbf{0}\}$ ,

**(iii)**  $\dim V + \dim V^\perp = n$ .

**Theorem 3** If  $V$  is the row space of a matrix, then  $V^\perp$  is the nullspace of the same matrix.



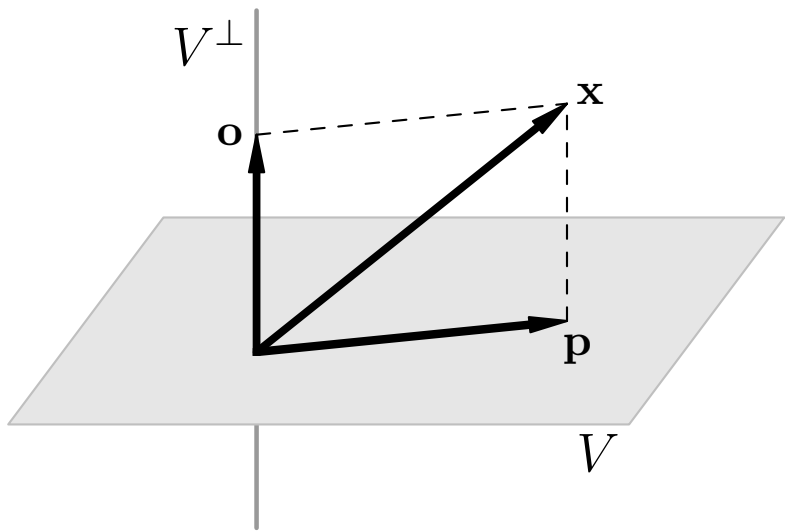
## Orthogonal projection

**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  $\mathbf{o} \in V^\perp$ .

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  $V$ .



**Problem.** Let  $\Pi$  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  $\Pi$ .

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

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  $\Pi$  is  $\mathbf{p}$  and the distance from  $\mathbf{x}$  to  $\Pi$  is  $\|\mathbf{o}\|$ .

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

Then  $\mathbf{o} = \mathbf{x} - \mathbf{p} = \mathbf{x} - \alpha\mathbf{v}_1 - \beta\mathbf{v}_2$ .

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

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

---

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

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

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

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

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



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:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \iff \mathbf{Ax} = \mathbf{b}$$

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

The **least squares solution**  $\mathbf{x}$  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**  $A^T A\mathbf{x} = A^T \mathbf{b}$ .

*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

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

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

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

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

**Problem.** Find the constant function that is the least square fit to the following data

$x$	0	1	2	3
$f(x)$	1	0	1	2

$$f(x) = c \implies \begin{cases} c = 1 \\ c = 0 \\ c = 1 \\ c = 2 \end{cases} \implies \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (c) = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$(1, 1, 1, 1) \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (c) = (1, 1, 1, 1) \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$c = \frac{1}{4}(1 + 0 + 1 + 2) = 1 \quad (\text{mean arithmetic value})$$

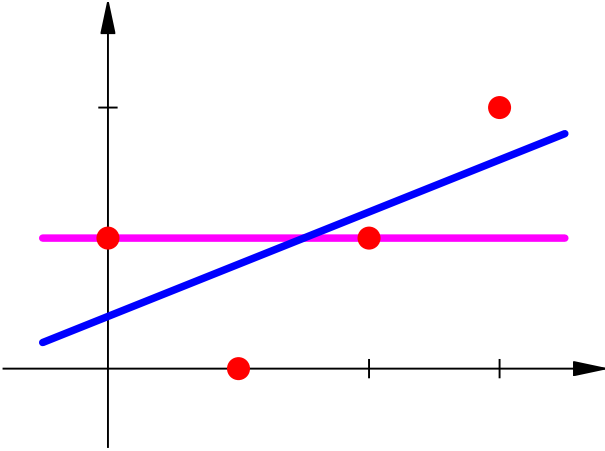
**Problem.** Find the linear polynomial that is the least square fit to the following data

$x$	0	1	2	3
$f(x)$	1	0	1	2

$$f(x) = c_1 + c_2x \implies \begin{cases} c_1 = 1 \\ c_1 + c_2 = 0 \\ c_1 + 2c_2 = 1 \\ c_1 + 3c_2 = 2 \end{cases} \implies \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \end{pmatrix} \iff \begin{cases} c_1 = 0.4 \\ c_2 = 0.4 \end{cases}$$



## Norm

The notion of *norm* generalizes the notion of length of a vector in  $\mathbb{R}^n$ .

*Definition.* Let  $V$  be a vector space. A function  $\alpha : V \rightarrow \mathbb{R}$  is called a **norm** on  $V$  if it has the following properties:

- (i)  $\alpha(\mathbf{x}) \geq 0$ ,  $\alpha(\mathbf{x}) = 0$  only for  $\mathbf{x} = \mathbf{0}$  (positivity)
- (ii)  $\alpha(r\mathbf{x}) = |r| \alpha(\mathbf{x})$  for all  $r \in \mathbb{R}$  (homogeneity)
- (iii)  $\alpha(\mathbf{x} + \mathbf{y}) \leq \alpha(\mathbf{x}) + \alpha(\mathbf{y})$  (triangle inequality)

*Notation.* The norm of a vector  $\mathbf{x} \in V$  is usually denoted  $\|\mathbf{x}\|$ . Different norms on  $V$  are distinguished by subscripts, e.g.,  $\|\mathbf{x}\|_1$  and  $\|\mathbf{x}\|_2$ .



*Examples.*  $V = \mathbb{R}^n$ ,  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ .

- $\|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|, \dots, |x_n|)$ .

Positivity and homogeneity are obvious.

The triangle inequality:

$$\begin{aligned} |x_i + y_i| &\leq |x_i| + |y_i| \leq \max_j |x_j| + \max_j |y_j| \\ \implies \max_j |x_j + y_j| &\leq \max_j |x_j| + \max_j |y_j| \end{aligned}$$

- $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \dots + |x_n|$ .

Positivity and homogeneity are obvious.

The triangle inequality:  $|x_i + y_i| \leq |x_i| + |y_i|$

$$\implies \sum_j |x_j + y_j| \leq \sum_j |x_j| + \sum_j |y_j|$$

*Examples.*  $V = \mathbb{R}^n$ ,  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ .

- $\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \dots + |x_n|^p)^{1/p}$ ,  $p > 0$ .

**Theorem**  $\|\mathbf{x}\|_p$  is a norm on  $\mathbb{R}^n$  for any  $p \geq 1$ .

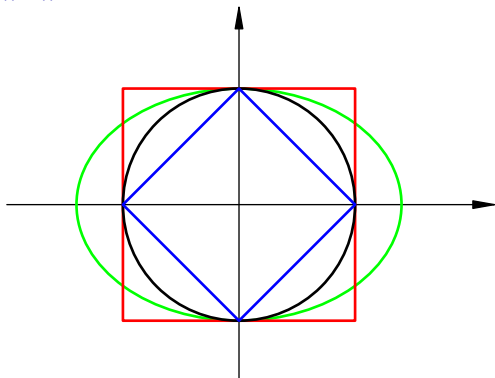
*Remark.*  $\|\mathbf{x}\|_2 =$  Euclidean length of  $\mathbf{x}$ .

*Definition.* A **normed vector space** is a vector space endowed with a norm.

The norm defines a distance function on the normed vector space:  $\text{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ .

Then we say that a sequence  $\mathbf{x}_1, \mathbf{x}_2, \dots$  converges to a vector  $\mathbf{x}$  if  $\text{dist}(\mathbf{x}, \mathbf{x}_n) \rightarrow 0$  as  $n \rightarrow \infty$ .

Unit circle:  $\|\mathbf{x}\| = 1$



$$\|\mathbf{x}\| = (x_1^2 + x_2^2)^{1/2} \quad \text{black}$$

$$\|\mathbf{x}\| = \left(\frac{1}{2}x_1^2 + x_2^2\right)^{1/2} \quad \text{green}$$

$$\|\mathbf{x}\| = |x_1| + |x_2| \quad \text{blue}$$

$$\|\mathbf{x}\| = \max(|x_1|, |x_2|) \quad \text{red}$$

*Examples.*  $V = C[a, b]$ ,  $f : [a, b] \rightarrow \mathbb{R}$ .

- $\|f\|_{\infty} = \max_{a \leq x \leq b} |f(x)|.$

- $\|f\|_1 = \int_a^b |f(x)| dx.$

- $\|f\|_p = \left( \int_a^b |f(x)|^p dx \right)^{1/p}, \quad p > 0.$

**Theorem**  $\|f\|_p$  is a norm on  $C[a, b]$  for any  $p \geq 1$ .