ORTHOGONALITY AND LEAST-SQUARES [CHAP. 6]

Inner products and Norms

➤ Inner product or dot product of 2 vectors u and v in $\mathbb{R}^n$:

$$u.v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

✍ Calculate $u.v$ when $u = \begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix} \qquad v = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 5 \end{bmatrix}$

➤ If u and v are vectors in $\mathbb{R}^n$ then we can regard u and v as $n \times 1$ matrices. The transpose u^T is a $1 \times n$ matrix, and the matrix product $u^T v$ is a 1×1 matrix $=$ a scalar.

➤ Then note that $\boxed{u.v = v.u = u^T v = v^T u}$

Length of a vector in $\mathbb{R}^n$

Euclidean norm of a vector u is $\|u\| = \sqrt{u.u}$, i.e.,

$$\|u\| = (u.u)^{1/2} = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}$$

- This is the length of vector u

- If we identify v with a geometric point in the plane, then $\|v\|$ is the standard notion of the length of the line segment from 0 to v.

- This follows from the Pythagorean Theorem applied to a triangle..

- A vector of length one is often called a unit vector

- The process of dividing a vector by its length to create a vector of unit length (a unit vector) is called normalizing

✍ Normalize $v = (1; -2; 2; 0)$.

Important properties

- For any scalar α, the length αv is $|\alpha|$ times the length of v. That is,

$$\|\alpha v\| = |\alpha| \|v\|$$

- The length of the sum of any two vectors does not exceed the sum of the lengths of the vectors (Triangle inequality)

$$\boxed{\|u + v\| \leq \|u\| + \|v\|}$$

- The Cauchy-Schwartz inequality :

$$\boxed{|x.y| \leq \|x\| \|y\|}$$

Distance in $\mathbb{R}^n$

Definition: The distance between $\boldsymbol{u}$ and $\boldsymbol{v}$, two vectors in $\mathbb{R}^n$ is the length of the vector $\boldsymbol{u} - \boldsymbol{v}$

➤ Written as $\boldsymbol{dist(u, v)}$ or $\boldsymbol{d(u, v)}$

$$\boldsymbol{d(u, v) = \|u - v\|}$$

✍ Distance between $\boldsymbol{u} = \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \end{pmatrix}$ and $\boldsymbol{v} = \begin{pmatrix} \mathbf{4} \\ \mathbf{-3} \end{pmatrix}$

Orthogonality

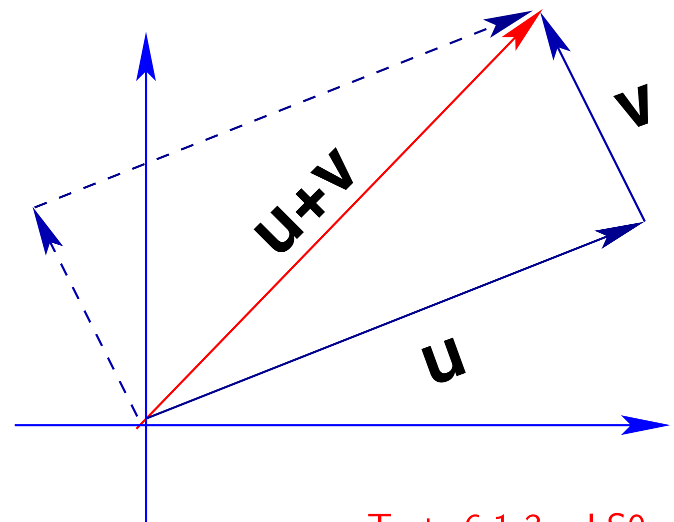
1. Two vectors u and v are orthogonal if $(u, v) = 0$.
2. A system of vectors $\{v_1, \ldots, v_n\}$ is orthogonal if $(v_i, v_j) = 0$ for $i \neq j$; and orthonormal if $(v_i, v_j) = \delta_{ij}$

Pythagoras theorem:

$$u \perp v \quad \Leftrightarrow \quad \|u + v\|^2 = \|u\|^2 + \|v\|^2$$

That is, two vectors u and v are orthogonal if and only if

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$



Least-Squares systems – Background

- Recall orthogonality: $x \perp y$ if $x.y = 0$
- Equivalently $x \perp y$ if $y^T x = 0$ or $x^T y = 0$
- A zero vector is trivially orthogonal to any vector.

- A vector x is orthogonal to a subspace S if:

$$\boxed{x \perp y \text{ for all } y \in S}$$

- If $A = [a_1, a_2, \cdots, a_n]$ is a basis of S then

$$x \perp S \quad \leftrightarrow \quad A^T x = 0 \quad \leftrightarrow \quad x^T A = 0$$

- The space of all vectors orthogonal to S is a subspace.

Notation: $S^{\perp}$

- Two subspaces S_1, S_2 are orthogonal to each other when

$$x \perp y \quad \text{for all } x \text{ in } S_1, \quad \text{for all } y \text{ in } S_2$$

✍ Show that

$$\text{Nul}(\boldsymbol{A}) \perp \text{Col}(\boldsymbol{A}^T) \quad \text{and}$$
$$\text{Nul}(\boldsymbol{A}^T) \perp \text{Col}(\boldsymbol{A})$$

➤ Indeed: $\boldsymbol{Ax = 0}$ means $(\boldsymbol{A}^T)^T\boldsymbol{x} = \boldsymbol{0}$. So if $\boldsymbol{x} \in \boldsymbol{Nul(A)}$, it is $\perp$ to the columns of $\boldsymbol{A}^T$, i.e., to the range of $\boldsymbol{A}^T$. Second result: replace $\boldsymbol{A}$ by $\boldsymbol{A}^T$.

✍ Find the subspace of all vectors that are orthogonal to $\mathbf{span}\{\boldsymbol{v_1, v_2}\}$ where

$$[\boldsymbol{v_1, v_2}] = \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{-1} & \mathbf{0} \\ \mathbf{1} & \mathbf{-1} \end{bmatrix}$$

Least-Squares systems

Problem: Given: an $m \times n$ matrix and a right-hand side b in $\mathbb{R}^m$, find $x \in \mathbb{R}^n$ which minimizes:

$$\|b - Ax\|$$

Assumption: $m > n$ and $\text{rank}(A) = n$ ('A is of full rank')

✍ Find equivalent conditions to this assumption

Theorem If A has full rank then $A^T A$ is invertible.

Proof We need to prove: $A^T Ax = 0$ implies $x = 0$.
Assume $A^T Ax = 0$. Then $x^T A^T Ax = 0$ – i.e., $(Ax)^T Ax = 0$, or $\|Ax\|^2 = 0$. This means $Ax = 0$. But since the columns of A are independent x must be zero. QED.

Theorem Let $\boldsymbol{A}$ be an $\boldsymbol{m} \times \boldsymbol{n}$ matrix of rank $\boldsymbol{n}$. Then $\boldsymbol{x}^*$ is the solution of the least-squares problem $\min \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|$

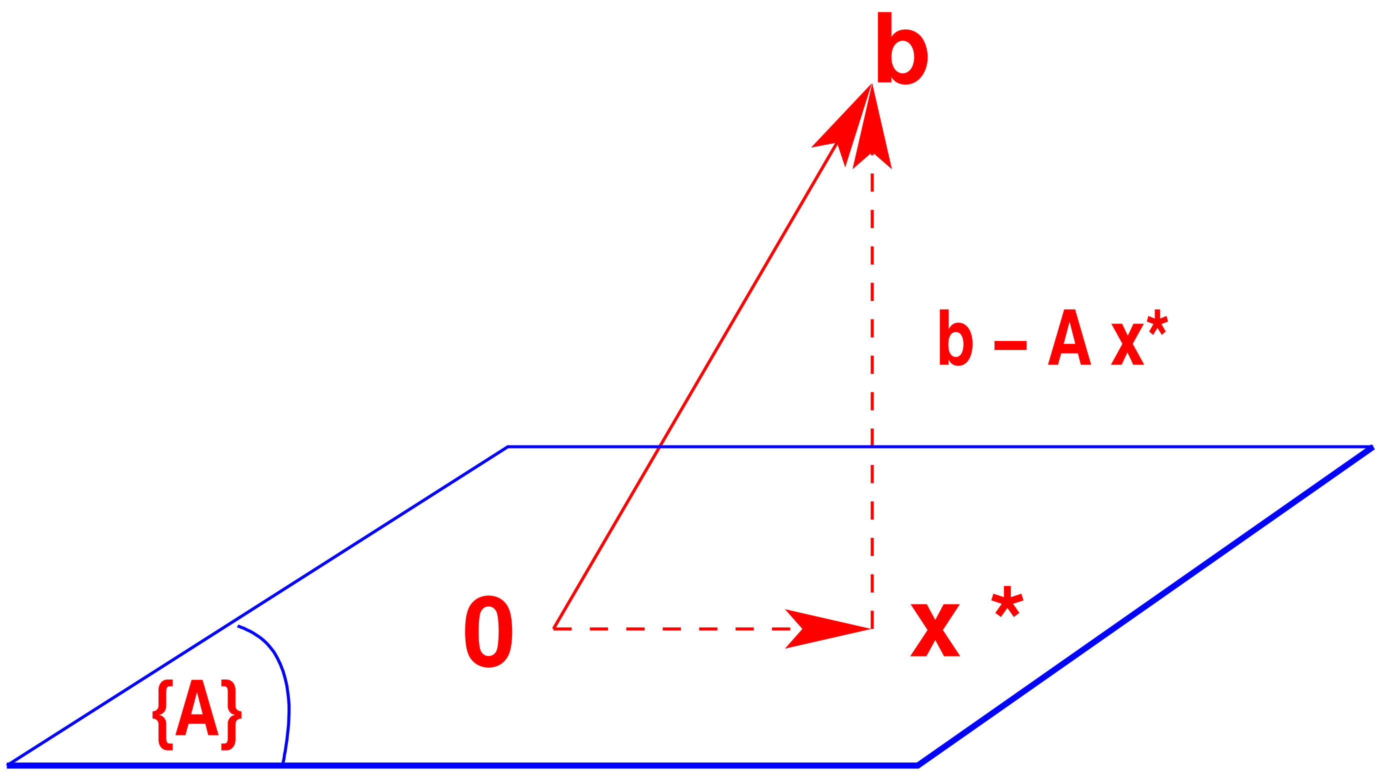
if and only if $\boxed{\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* \perp \mathrm{Col}(\boldsymbol{A})}$

if and only if $\boxed{\boldsymbol{A}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*) = \boldsymbol{0}}$

if and only if $\boxed{\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}^T\boldsymbol{b}}$

Proof See text.

Illustration of theorem: $\boldsymbol{x}^*$ is the best approximation to the vector $\boldsymbol{b}$ from the subspace $\mathbf{span}\{\boldsymbol{A}\}$ if and only if $\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}^*$ is $\perp$ to the whole subspace $\mathbf{span}\{\boldsymbol{A}\}$. This in turn is equivalent to $\boldsymbol{A}^T(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}^*)=\mathbf{0}$ ➤ $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}^T\boldsymbol{b}$. Note: $\mathbf{span}\{\boldsymbol{A}\}=$ Col($\boldsymbol{A}$) $=$ column space of $\boldsymbol{A}$



Normal equations

➤ The system

$$\boxed{A^T A x = A^T b}$$

is called the system of normal equations for the matrix A and rhs b

➤ Its solution is the solution of the least-squares problem $\min \|b - Ax\|$

✍ Find the least solution by solving the normal equations when:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & -1 & 1 \\ 1 & 1 & -2 \\ 0 & 2 & 1 \end{bmatrix} \qquad b = \begin{bmatrix} 2 \\ 0 \\ 4 \\ 1 \end{bmatrix}$$

Application: Linear data fitting

- Experimental data (not accurate) provides measurements $y_1, \ldots, y_m$ of an unknown linear function ϕ at points $t_1, \ldots, t_m$. Problem: find the 'best' possible approximation to ϕ.

- Must find:

$$\phi(t) = \beta_0 + \beta_1 t \ \text{ s.t. } \quad \phi(t_j) \approx y_j, j = 1, \ldots, m$$

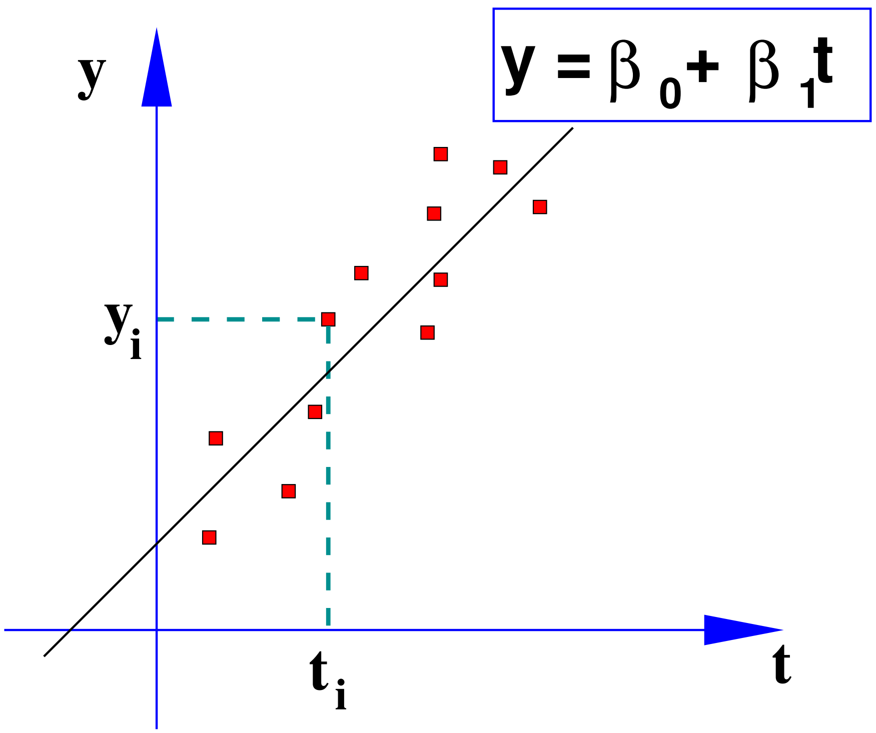
- Question: Close in what sense?

- Least-squares approximation sense: Find ϕ such that

$$|\phi(t_1) - y_1|^2 + |\phi(t_2) - y_2|^2 + \cdots + |\phi(t_m) - y_m|^2 = \text{Min}$$

➤ We want to find best fit in least-squares sense for the equations

$$\begin{aligned} \beta_0 + \beta_1 t_1 &= y_1 \\ \beta_0 + \beta_1 t_2 &= y_2 \\ \vdots \quad &= \vdots \\ \beta_0 + \beta_1 t_m &= y_m \end{aligned}$$



➤ Using matrix notation this means: find 'best' approximation to vector y from linear combinations of vectors f_1, f_2, where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, \quad f_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad f_2 = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{pmatrix}$$

- Define

$$F = [f_1, f_2], \quad x = \binom{\beta_0}{\beta_1}$$

- We want to find x such $\|Fx - y\|$ is minimum.
- Least-squares linear system. F is $m \times 2$.

The vector x_* mininizes $\|y - Fx\|$ if and only if it is the solution of the **normal equations:**

$$F^T F x = F^T y$$