ORTHOGONALITY AND LEAST-SQUARES [CHAP. 6]

Inner products and Norms

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

$$u.v = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

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

➤ 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^Tv is a 1×1 matrix $=$ a scalar.

➤ Then note that $\quad u.v = v.u = u^Tv = v^Tu$

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

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

$$\|\boldsymbol{u}\| = (\boldsymbol{u}.\boldsymbol{u})^{1/2} = \sqrt{\boldsymbol{u}_1^2 + \boldsymbol{u}_2^2 + \ldots + \boldsymbol{u}_n^2}$$

- This is the length of vector $\boldsymbol{u}$
- If we identify $\boldsymbol{v}$ with a geometric point in the plane, then $\|\boldsymbol{v}\|$ is the standard notion of the length of the line segment from 0 to $\boldsymbol{v}$.
- 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 $\boldsymbol{v} = [\mathbf{1}; -\mathbf{2}; \mathbf{2}; \mathbf{0}]$. [Matlab notation used]

Important properties

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

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

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

$$\|u + v\| \leq \|u\| + \|v\|$$

- The Cauchy-Schwartz inequality :

$$|u.v| \leq \|u\| \|v\|$$

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 $\boldsymbol{v} = \begin{pmatrix} \mathbf{4} \\ \mathbf{-3} \end{pmatrix}$

➤ See illustration in Example 4 of *text* .

Orthogonality

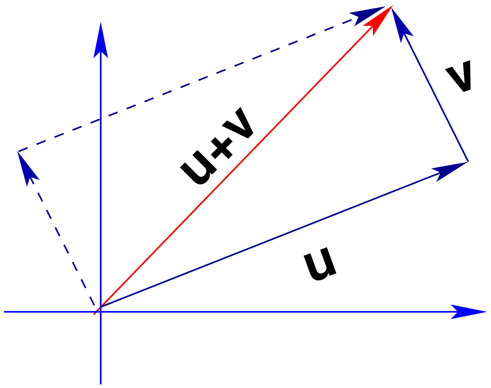
1. Two vectors u and v are orthogonal if $u.v = 0$.
   Common notation: $u \perp v$
2. A system of vectors $S = \{v_1, \ldots, v_n\}$ is orthogonal if $v_i.v_j = 0$ for $i \neq j$.

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

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



Orthogonal systems (continued)

✍ Show that the following system is orthogonal

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \quad v_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad v_3 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

Theorem: If $S = \{v_1, ..., v_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent. Hence S is a basis of $\mathbf{span}(S)$.

Definition: An orthogonal basis of a subspace W is a basis of W that is also an orthogonal set.

✍ Let $S = \{v_1, ..., v_p\}$ be an orthogonal basis of a subspace W. Then a vector x in W is a linear combination of the v_i's:

$$x = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_p v_p$$

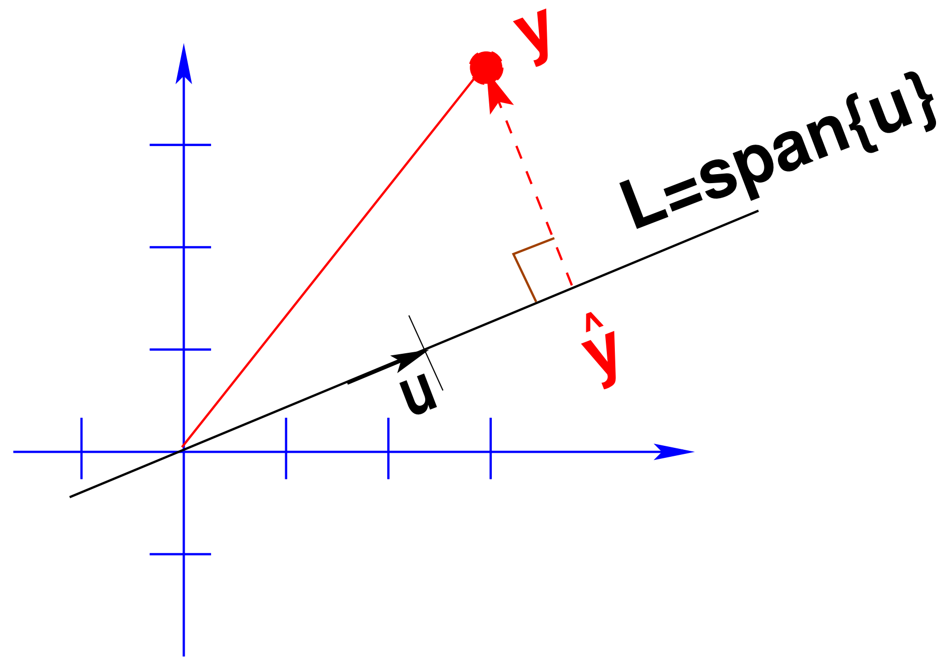
How can you get the α_i's? [Hint: Compute the inner product of x with each v_i.]

✍ Read Section 6.2 of *text* – specifically the paragraph on orthogonal projection (p. 342) for a geometric interpretation.

➤ We say that a system of vectors $S = \{v_1, ..., v_p\}$ is orthonormal if it is orthogonal and in addition each v_i has unit length, i.e., $\|v_i\| = 1$.

A brief introduction to least-squares

➤ Consider the following problem: find a member of the subspace $\boldsymbol{L} = \mathbf{span}\{\boldsymbol{u}\}$ that is closest to a vector $\boldsymbol{y}$ that does not belong to $\boldsymbol{L}$. How would you solve this geometrically?



➤ The solution $\hat{\boldsymbol{y}}$ is best approximatiob of $\boldsymbol{y}$ from $\boldsymbol{L}$

Answer: The line joining $\boldsymbol{y}$ to the best approximation $\hat{\boldsymbol{y}}$ should be orthogonal to $\boldsymbol{u}$:

$$\boldsymbol{y} - \hat{\boldsymbol{y}} \perp \boldsymbol{u}$$

- Since Write $\hat{u}$ is in L, we can write $\hat{u} = \alpha u$.
- Expand the orthogonality condition: $\boxed{u.y - u.(\alpha u) = 0} \rightarrow$

$$\alpha = \frac{u.y}{u.u}$$

✍ Solve the problem when $u = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ $\quad y = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and provide a geometric illustration.

- See Example 3 in Section 6.2 of *text*.

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

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

✍ 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: $\boldsymbol{Ax} = \boldsymbol{0}$ means $(\boldsymbol{A}^T)^T\boldsymbol{x} = \boldsymbol{0}$. So if $\boldsymbol{x} \in \boldsymbol{Nul}(\boldsymbol{A})$, 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 $\mathbf{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ where

$$[\boldsymbol{v}_1, \boldsymbol{v}_2] = \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{-1} & \mathbf{0} \\ \mathbf{1} & \mathbf{-1} \end{bmatrix}$$

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: $A^T A x = 0$ implies $x = 0$.
Assume $A^T A x = 0$. Then $x^T A^T A x = 0$ – 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 $\boldsymbol{m \times n}$ matrix of rank $\boldsymbol{n}$. Then $\boldsymbol{x^*}$ is the solution of the least-squares problem $\mathbf{min}\, \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|$

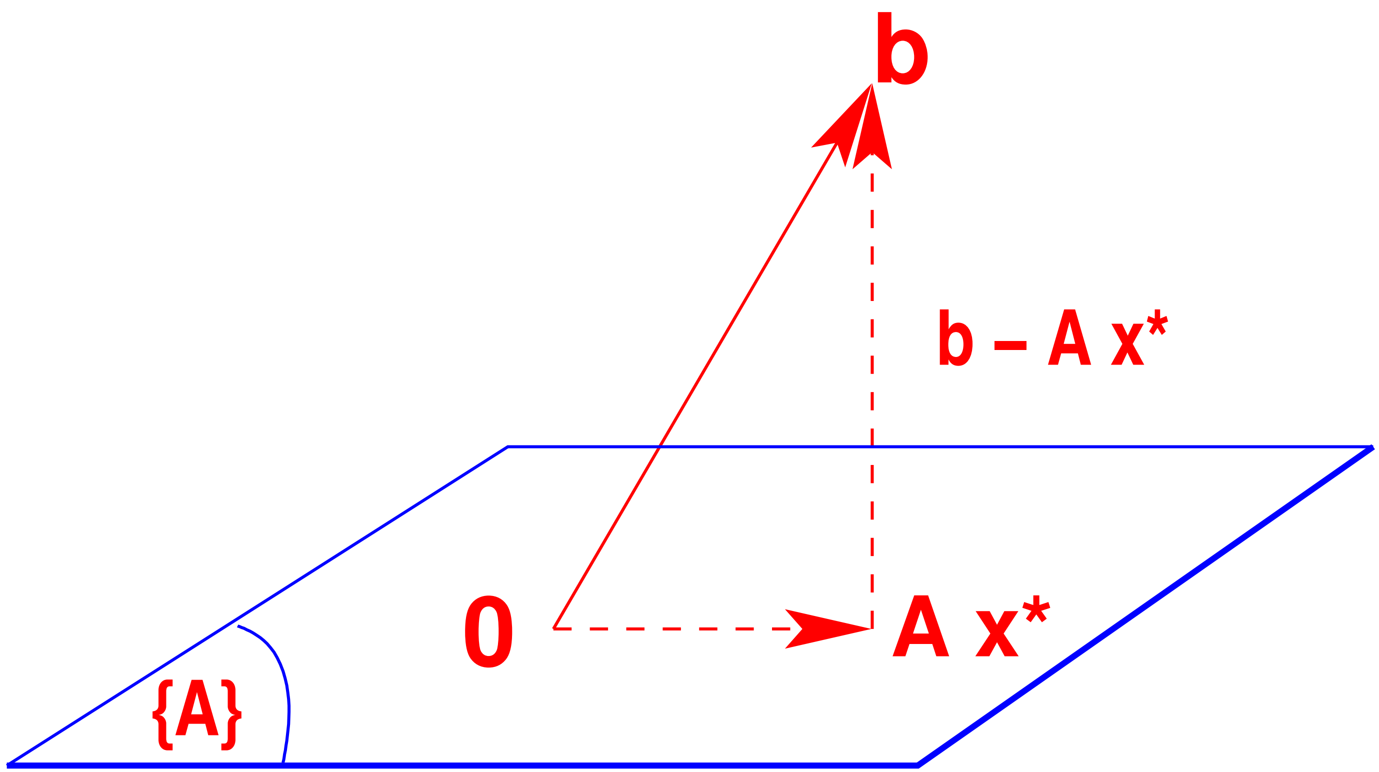
if and only if $\boxed{\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* \perp \mathrm{Col}(\boldsymbol{A})}$

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

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

Proof See text.

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



Normal equations

- The system

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

is called the system of normal equations for the matrix $\boldsymbol{A}$ and rhs $\boldsymbol{b}$

- Its solution is the solution of the least-squares problem $\min \|\boldsymbol{b} - \boldsymbol{A x}\|$

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

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

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:

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

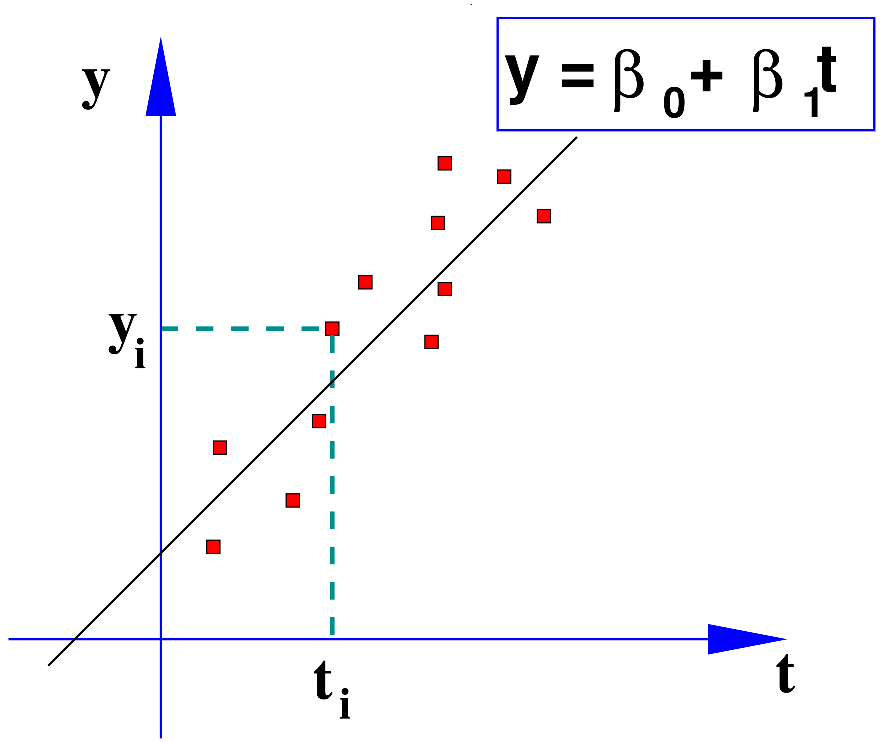
➤ Question: Close in what sense?

➤ 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

$$\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2], \quad \boldsymbol{x} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

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

The vector $\boldsymbol{x}_*$ mininizes $\|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{x}\|$ if and only if it is the solution of the **normal equations:**

$$\boldsymbol{F}^T\boldsymbol{F}\boldsymbol{x} = \boldsymbol{F}^T\boldsymbol{y}$$