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 $u.v = v.u = u^Tv = v^Tu$

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 $v = [1; -2; 2; 0]$. [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)

$$\|u + v\| \le \|u\| + \|v\|$$

➤ The Cauchy-Schwartz inequality :

$$|u.v| \le \|u\| \|v\|$$

Distance in $\mathbb{R}^n$

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

- Written as $dist(u, v)$ or $d(u, v)$

$$d(u, v) = \|u - v\|$$

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

- See illustration in Example 4 of *text* .

Orthogonality

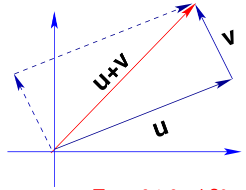
1. Two vectors u and v are orthogonal if $u.v = 0$.
   Common notation: $u \perp v$
2. A system of vectors $S = \{v_1, \ldots, v_n\}$ is orthogonal if $v_i.v_j = 0$ for $i \neq j$.

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

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

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

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

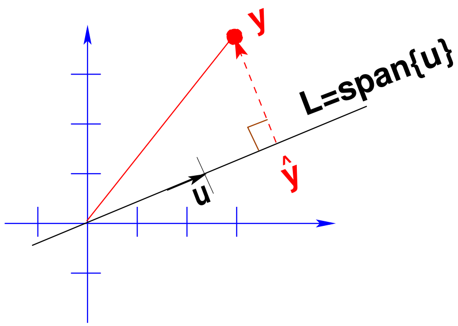
How can you get the α_i's? [Hint: Compute the inner product of x with each v_i.]

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 $S = \{v_1, \ldots, v_p\}$ 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 $L = \text{span}\{u\}$ that is closest to a vector y that does not belong to L. How would you solve this geometrically?



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

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

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

- 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 $u = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ $y = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ 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

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

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

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

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

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., $(Ax)^T A x = 0$, or $\|Ax\|^2 = 0$. This means $Ax = 0$. But since the columns of A are independent x must be zero. QED.

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

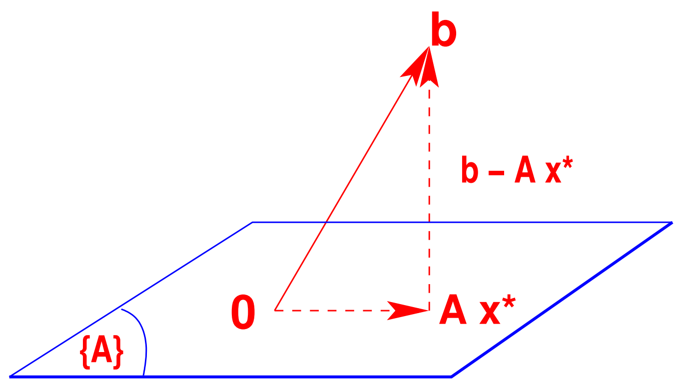
if and only if $\boxed{b - Ax^* \perp \text{Col}(A)}$

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

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

Proof See text.

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



Normal equations

- The system

$$A^T A x = A^T b$$

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:

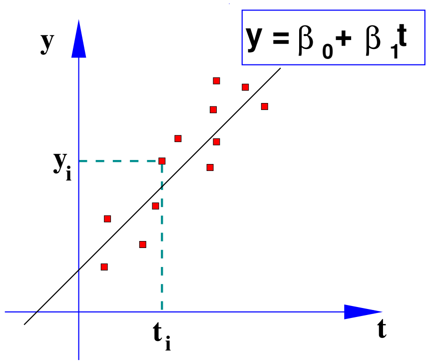
$$\phi(t) = \beta_0 + \beta_1 t \;\text{ s.t. }\; \phi(t_j) \approx y_j, j = 1, \ldots, m$$

- 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

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



- 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

$$F = [f_1, f_2], \quad x = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

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

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

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