ICS 6N Computational Linear Algebra
Orthogonality and Least Squares

Xiaohui Xie

University of California, Irvine

xhx@uci.edu

Inner product

- Let x, y be vectors in R^n. The inner product between x and y is defined to be

$$\begin{aligned} x \cdot y &= x_1 y_1 + x_2 y_2 + \ldots + x_n y_n \\ &= x^T y \end{aligned}$$

- x, y are called **orthogonal** if $x \cdot y = 0$

Norm

- The length (or norm) of a vector R^n is defined by

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x^T x}$$

- Provide a measure on the length of a vector
- Satisfying the following three properties:
  - $\|x\| \geq 0$
  - Triangular inequality: $\|y + x\| \leq \|x\| + \|y\|$
  - $\|cx\| = |c|\|x\|$
- A vector whose length is 1 is called a unit vector.

Distance in R^n

- For u and v in R^n, the distance between u and v, written as dist(u,v), is the length of the vector $u - v$. That is

$$\text{dist}(u, v) = \|u - v\|$$

Distance in R^n: an example

- Compute the distance between vectors $u = (7, 1)$ and $v = (3, 2)$.
- Calculate
$$u - v = \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$
- $$\|u - v\| = \sqrt{4^2 + (-1)^2} = \sqrt{17}$$

If $x \cdot y = 0$, what is $\|y - x\|$?

Solution

$$\|y - x\|^2 = (y - x)^T (y - x)$$
$$= (y^T - x^T)(y - x)$$
$$= y^T y - y^T x - x^T y + x^T x^T$$
$$= \|y\|^2 - 2x^T y + \|x\|^2$$

And since $x \cdot y = x^T y = 0$ we have

$$\|y - x\|^2 = \|y\|^2 + \|x\|^2$$

Orthogonal set

- A set of nonzero vectors $\{u_1, u_2, \ldots, u_p\}$ in R^n is said to be **orthogonal** if $u_i \cdot u_j = 0$ for any $i \neq j$
- The set is **orthonormal** if it is orthogonal and $\|u_i\| = 1$ for $i = 1, 2, \ldots, p$

Example

$$u_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, u_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, u_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

We can check: $u_1^T u_2 = 0$ $u_1^T u_3 = 0$ $u_2^T u_3 = 0$. So it is an orthogonal set.

If $W = span\{u_1, \dots, u_p\}$, then $\{u_1, \dots, u_p\}$ forms a basis for W, called **orthogonal basis**.

Orthogonal basis

- Let $W = span\{u_1, \ldots, u_p\}$ where $\{u_1, \ldots, u_p\}$ is an orthogonal basis of W.
- Let x be a vector in W. Find out $c_1, c_2, \ldots, c_p$ such that

$$x = c_1 u_1 + c_2 u_2 + \ldots + c_p u_p$$

- Since $u_i^T x = u_i^T (c_1 u_1 + c_2 u_2 + \ldots + c_p u_p) = c_i u_i^T u_i$,

$$c_i = \frac{u_i^T x}{u_i^T u_i}$$

Matrix with orthonormal columns

Suppose $\{u_1, u_2, \dots, u_n\}$ is an orthonormal set in R^m
Let U be an mxn matrix defined as:

$$U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$$

Then

$$U^T U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} = I_n$$

Because the entries are orthogonal

Properties of matrices with orthonormal columns

a) $\|Ux\| = \|x\|$ for any x in R^n
b) $(U \cdot x)(U \cdot y) = x \cdot y$ for any x, y in R^n
c) $Ux \cdot Uy = 0$ if $x \cdot y = 0$

Proof:
a) $\|Ux\|^2 = (Ux)^T Ux = x^T U^T Ux = x^T x = \|x\|^2$
b) $Ux \cdot Uy = (Ux)^T Uy = x^T U^T Uy = x^T y = x \cdot y$

Orthogonal matrix

- An nxn matrix U is called an **orthogonal matrix** if its column vectors are orthonormal.
- $U^T U = I$
- $U^{-1} = U^T$

Orthogonal projection

Given a vector u in R^n, consider the problem of decomposing a vector y in R^n into two components:

$$y = \hat{y} + z$$

where $\hat{y}$ is in $span\{u\}$ and z is orthogonal to u. $\hat{y}$ is called the *orthogonal projection* of y onto u.

Orthogonal projection

- Decompose $y = \hat{y} + z$
- Let $\hat{y} = \alpha u$ for some scalar α. Then

$$z = y - \hat{y} = y - \alpha u$$

- Since $z \cdot u = 0$, then

$$u^T(y - \alpha u) = 0$$

so we have $u^T y = \alpha u^T u$, and

$$\alpha = \frac{y^T u}{u^T u}$$

- $\mathrm{Proj}_u(y) = \hat{y} = \alpha u = \frac{y^T u}{u^T u} u$

Orthogonal projection: an example

Let $y = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $u = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Find the orthogonal projection of y onto u.

Orthogonal projection

Let $W = span\{u_1, \ldots, u_p\}$ is a subspace of R^n, where $\{u_1, \ldots, u_p\}$ is an orthogonal set. Decompose y into two components:

$$y = \hat{y} + z$$

where $\hat{y}$ is a vector in W and z is orthogonal to W. $\hat{y}$ is called the orthogonal projection of y onto W.

- Since $\hat{y}$ is in W, write

$$\hat{y} = c_1 u_1 + c_2 u_2 + \ldots + c_p u_p$$

- $z = y - \hat{y}$ is orthogonal to W, implying that $u_i \cdot z = 0$ for every i.
- From $u_i^T (y - \hat{y}) = 0 \implies c_i = \frac{u_i^T y}{u_i^T u_i}$
- So the orthogonal project of y onto W is

$$\mathrm{Proj}_W(y) = \hat{y} = \frac{u_1^T y}{u_1^T u_1} u_1 + \ldots + \frac{u_p^T y}{u_p^T u_p} u_p$$

Orthogonal projection: an example

Let $u_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $u_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Find the orthogonal projection of y onto $W = \mathrm{Span}\{u_1, u_2\}$.

Best approximation theorem

- Let W be a subspace of R^n, and let y be any vector in R^n. Let $\hat{y}$ be the orthogonal projection of y onto W. Then $\hat{y}$ is the **closest point** in W to y, in the sense that

$$\|y - \hat{y}\| < \|y - v\|$$

for all v in W distinct from $\hat{y}$.

Best approximation theorem

- Let W be a subspace of R^n, and let y be any vector in R^n. Let $\hat{y}$ be the orthogonal projection of y onto W. Then $\hat{y}$ is the **closest point** in W to y, in the sense that

  $$\|y - \hat{y}\| < \|y - v\|$$

  for all v in W distinct from $\hat{y}$.
- PROOF: Take v in W distinct from $\hat{y}$. Then

  $$\begin{aligned}\|y - v\| &= \|y - \hat{y} + \hat{y} - v\| \\ &= \|y - \hat{y}\| + \|\hat{y} - v\| > \|y - \hat{y}\|\end{aligned}$$

Finding orthogonal basis

Given a basis $\{x_1, \ldots, x_p\}$ for a subspace W of R^n, find an orthogonal basis $\{v_1, \ldots, v_p\}$ for W such that for any $i = 1, \cdots, p$

$$span\{v_1, \cdots, v_i\} = span\{x_1, \cdots, x_i\}$$

Gram-Schmidt process for finding orthogonal basis

Given a basis $\{x_1, \ldots, x_p\}$ for a subspace W of R^n, find an orthogonal basis $\{v_1, \ldots, v_p\}$ for W such that for any $i = 1, \cdots, p$

$$span\{v_1, \cdots, v_i\} = span\{x_1, \cdots, x_i\}$$

1. $v_1 = x_1$
2. $v_2 = x_2 - Proj_{v_1}(x_2) = x_2 - \frac{x_2^T v_1}{v_1^T v_1} v_1$
3. $v_3 = x_3 - Proj_{span\{v_1, v_2\}}(x_3) = x_3 - \frac{x_3^T v_1}{v_1^T v_1} v_1 - \frac{x_3^T v_2}{v_2^T v_2} v_2$

   $\vdots$
4. $v_p = x_p - Proj_{span\{v_1, \ldots v_{p-1}\}}(x_p)$
   $= x_p - \frac{x_p^T v_1}{v_1^T v_1} v_1 - \frac{x_p^T v_2}{v_2^T v_2} v_2 + \ldots + \frac{x_p^T v_{p-1}}{v_{p-1}^T v_{p-1}} v_{p-1}$

Gram-Schmidt process: an example

Let $x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and $x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Construct an orthogonal basis for $W = \mathrm{Span}\{x_1, x_2, x_3\}$.

Least square problems

- Suppose $Ax = b$ has no solutions. Can we still find a solution x such that Ax is "closest" to b?
- Most common cases: A is an mxn matrix with $m > n$. The system $Ax = b$ has more equations than variables. So in general there is no solution.
- "Best solution" in the following sense: Find $\hat{x}$ such that $A\hat{x}$ is the **closest point** to b. That is,

$$\|A\hat{x} - b\| \leq \|Ax - b\|$$

  for all x in R^n.
- $\hat{x}$ is called the **least square solution**.

Find least square solutions of $Ax = b$

- Problem: Find $\hat{x}$ such that $A\hat{x}$ is closest to b.
- The problem is equivalent to finding a point $\hat{b}$ in Col A that is closest to b.
- From the best approximation theorem, the point in Col A closest to b is the orthogonal projection of b onto Col A:

$$A\hat{x} = \hat{b} = \mathrm{proj}_{Col\ A}\ b$$

Find least square solutions of $Ax = b$

- The point in Col A closest to b is

$$A\hat{x} = \mathrm{proj}_{Col\ A}\ b$$

- The residual $r = b - A\hat{x}$ is orthogonal to Col A, implying that

$$(b - A\hat{x}) \perp a_i \implies a_i^T(b - A\hat{x}) = 0 \text{ for all } i$$

  Written in matrix format, $A^T(b - A\hat{x}) = 0$

- So we have the normal equation

$$A^TA\hat{x} = A^Tb$$

- If A^TA is invertible, then

$$\hat{x} = (A^TA)^{-1}A^Tb$$

Least square problems: an example

Find a least-squares solution of the inconsistent system $Ax = b$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

- Compute

$$A^T A = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}, \quad A^T b = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

- Solve the normal equation $A^T A\hat{x} = A^T b$ using Gaussian elimination,

$$\hat{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$