

Lecture 13: Least Squares

1 Bridge to Least Squares

Let $L \subseteq \mathbb{R}^m$ be a linear subspace, and let $P_L : \mathbb{R}^m \rightarrow L$ denote the orthogonal projection onto L . Recall that for every $b \in \mathbb{R}^m$,

$$b = P_L b + (b - P_L b),$$

where

$$P_L b \in L, \quad b - P_L b \in L^\perp.$$

The orthogonal projection $P_L b$ is the point in L that is closest to b .

This simple geometric fact is the key idea behind least squares.

Suppose that

$$A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m,$$

and we want to solve

$$Ax = b.$$

Every vector of the form Ax belongs to the column space

$$\mathcal{C}(A) = \{Ax : x \in \mathbb{R}^n\}.$$

Therefore, solving $Ax = b$ means finding a point in $\mathcal{C}(A)$ that is exactly equal to b .

If

$$b \notin \mathcal{C}(A),$$

then the system has no exact solution. However, we can still ask:

Which vector $Ax \in \mathcal{C}(A)$ is closest to b ?

By the best approximation property, the answer must be

$$P_{\mathcal{C}(A)} b.$$

Thus, instead of solving

$$Ax = b,$$

we try to solve

$$Ax = P_{\mathcal{C}(A)} b.$$

This is the geometric foundation of least squares.

2 The Least-Squares Problem

Definition (Least-Squares Problem). *Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. A least-squares solution is a vector $\hat{x} \in \mathbb{R}^n$ satisfying*

$$\hat{x} \in \arg \min_{x \in \mathbb{R}^n} \|b - Ax\|_2^2.$$

The vector

$$r = b - A\hat{x}$$

is called the *residual*.

There are two cases.

2.1 Case 1: The System is Consistent

If

$$b \in \mathcal{C}(A),$$

then there exists x such that

$$Ax = b.$$

Consequently,

$$\|b - Ax\|_2^2 = 0,$$

which is clearly the smallest possible value.

Thus, whenever $Ax = b$ has an exact solution, every exact solution is also a least-squares solution.

2.2 Case 2: The System is Inconsistent

Suppose

$$b \notin \mathcal{C}(A).$$

Then

$$Ax = b$$

has no solution.

Nevertheless, since $Ax \in \mathcal{C}(A)$ for every x , minimizing

$$\|b - Ax\|_2$$

is exactly the problem of finding the point in $\mathcal{C}(A)$ closest to b .

Therefore,

$$A\hat{x} = P_{\mathcal{C}(A)}b.$$

Theorem 2.1 (Geometric Characterization of Least Squares). *A vector $\hat{x} \in \mathbb{R}^n$ is a least-squares solution of*

$$Ax \approx b$$

if and only if

$$A\hat{x} = P_{\mathcal{C}(A)}b.$$

Equivalently, the residual satisfies

$$b - A\hat{x} \in \mathcal{C}(A)^\perp.$$

Thus,

$$\boxed{b - A\hat{x} \perp \mathcal{C}(A).}$$

This orthogonality condition will give us another useful characterization of least-squares solutions.

3 Least Squares and the Normal Equations

Let

$$A = \begin{pmatrix} | & | & \cdots & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{pmatrix}.$$

The condition

$$b - A\hat{x} \perp \mathcal{C}(A)$$

means that the residual is orthogonal to every column of A :

$$a_j^T(b - A\hat{x}) = 0, \quad j = 1, \dots, n.$$

Combining these equations gives

$$A^T(b - A\hat{x}) = 0.$$

Therefore,

$$A^T A\hat{x} = A^T b.$$

Theorem 3.1 (Normal Equations). *A vector $\hat{x} \in \mathbb{R}^n$ is a least-squares solution of*

$$Ax \approx b$$

if and only if

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

Proof. We have

$\hat{x}$ is a least-squares solution

if and only if

$$b - A\hat{x} \perp \mathcal{C}(A).$$

This is equivalent to saying that $b - A\hat{x}$ is orthogonal to every column of A , which is equivalent to

$$A^T(b - A\hat{x}) = 0.$$

Rearranging gives

$$A^T A\hat{x} = A^T b.$$

□

The normal equations are mathematically convenient, but later we will see that solving least-squares problems using QR factorization is often numerically preferable to explicitly forming $A^T A$.

4 Solving Least Squares with QR

Suppose

$$A \in \mathbb{R}^{m \times n}$$

has linearly independent columns. Then A has a reduced QR factorization

$$A = QR,$$

where

$$Q \in \mathbb{R}^{m \times n}, \quad Q^T Q = I_n,$$

and

$$R \in \mathbb{R}^{n \times n}$$

is upper triangular and invertible.

Since the columns of Q form an orthonormal basis for $\mathcal{C}(A)$,

$$P_{\mathcal{C}(A)} = QQ^T.$$

A least-squares solution therefore satisfies

$$A\hat{x} = P_{\mathcal{C}(A)}b,$$

so that

$$QR\hat{x} = QQ^Tb.$$

Multiplying both sides by Q^T gives

$$Q^T QR\hat{x} = Q^T QQ^Tb.$$

Since

$$Q^T Q = I_n,$$

we obtain

$$\boxed{R\hat{x} = Q^Tb.}$$

Because R is upper triangular, this system can be solved efficiently using back substitution.

Theorem 4.1 (Least Squares via QR). *Suppose $A \in \mathbb{R}^{m \times n}$ has linearly independent columns and*

$$A = QR$$

is its reduced QR factorization. Then the unique least-squares solution of

$$Ax \approx b$$

is obtained by solving

$$\boxed{R\hat{x} = Q^Tb.}$$

5 Uniqueness of Least-Squares Solutions

The vector

$$A\hat{x} = P_{\mathcal{C}(A)}b$$

is always unique, because orthogonal projection onto a subspace is unique.

However, the vector $\hat{x}$ itself need not be unique.

Suppose $\hat{x}$ is a least-squares solution and

$$z \in \mathcal{N}(A).$$

Then

$$Az = 0,$$

and therefore

$$A(\hat{x} + z) = A\hat{x} + Az = A\hat{x}.$$

Thus,

$$\hat{x} + z$$

is another least-squares solution.

Theorem 5.1. *If $\hat{x}$ is one least-squares solution of $Ax \approx b$, then the set of all least-squares solutions is*

$$\boxed{\hat{x} + \mathcal{N}(A)}.$$

In particular, the least-squares solution is unique if and only if

$$\mathcal{N}(A) = \{0\}.$$

Equivalently, A must have linearly independent columns.

6 The Four Fundamental Subspaces

Let

$$A \in \mathbb{R}^{m \times n}.$$

Associated with A are four important linear subspaces.

Definition (Four Fundamental Subspaces). *The four fundamental subspaces of A are:*

1. *The column space*

$$\mathcal{C}(A) = \{Ax : x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m.$$

2. *The null space*

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\} \subseteq \mathbb{R}^n.$$

3. *The row space*

$$\mathcal{C}(A^T) \subseteq \mathbb{R}^n.$$

4. The left null space

$$\mathcal{N}(A^T) = \{y \in \mathbb{R}^m : A^T y = 0\} \subseteq \mathbb{R}^m.$$

Notice that the four spaces naturally come in two pairs:

$$\mathcal{C}(A^T), \mathcal{N}(A) \subseteq \mathbb{R}^n,$$

and

$$\mathcal{C}(A), \mathcal{N}(A^T) \subseteq \mathbb{R}^m.$$

7 Orthogonality of the Fundamental Subspaces

The four fundamental subspaces are closely related through orthogonality.

Theorem 7.1. *Let $A \in \mathbb{R}^{m \times n}$. Then*

$$\boxed{\mathcal{N}(A) = \mathcal{C}(A^T)^\perp}$$

and therefore

$$\boxed{\mathcal{C}(A^T) = \mathcal{N}(A)^\perp}.$$

Proof. Let $x \in \mathcal{N}(A)$. Then

$$Ax = 0.$$

For every $y \in \mathbb{R}^m$,

$$\langle A^T y, x \rangle = \langle y, Ax \rangle = 0.$$

Since every vector in $\mathcal{C}(A^T)$ has the form $A^T y$, we conclude that

$$x \perp \mathcal{C}(A^T).$$

Thus,

$$\mathcal{N}(A) \subseteq \mathcal{C}(A^T)^\perp.$$

Conversely, suppose

$$x \in \mathcal{C}(A^T)^\perp.$$

Then

$$\langle A^T y, x \rangle = 0$$

for every $y \in \mathbb{R}^m$. Hence

$$\langle y, Ax \rangle = 0$$

for every $y \in \mathbb{R}^m$.

Taking $y = Ax$ gives

$$\|Ax\|_2^2 = 0,$$

and therefore

$$Ax = 0.$$

Thus,

$$x \in \mathcal{N}(A).$$

□

Similarly,

Theorem 7.2. *Let $A \in \mathbb{R}^{m \times n}$. Then*

$$\mathcal{N}(A^T) = \mathcal{C}(A)^\perp$$

and

$$\mathcal{C}(A) = \mathcal{N}(A^T)^\perp.$$

Therefore,

$$\mathbb{R}^n = \mathcal{C}(A^T) \oplus \mathcal{N}(A)$$

and

$$\mathbb{R}^m = \mathcal{C}(A) \oplus \mathcal{N}(A^T).$$

Theorem 7.3. *Let $A \in \mathbb{R}^{m \times n}$. Then*

$$\text{rank}(A) = \text{rank}(A^T).$$

Proof. Let

$$A = QR$$

be a reduced QR factorization of A , where

$$Q \in \mathbb{R}^{m \times r}, \quad R \in \mathbb{R}^{r \times n},$$

and

$$r = \text{rank}(A).$$

Then

$$A^T = R^T Q^T.$$

We claim that

$$\mathcal{C}(A^T) = \mathcal{C}(R^T).$$

First, since

$$A^T y = R^T Q^T y,$$

every vector in $\mathcal{C}(A^T)$ belongs to $\mathcal{C}(R^T)$. Hence,

$$\mathcal{C}(A^T) \subseteq \mathcal{C}(R^T).$$

Conversely, let $z \in \mathcal{C}(R^T)$. Then

$$z = R^T w$$

for some $w \in \mathbb{R}^r$.

Since Q has orthonormal columns,

$$Q^T Q = I_r.$$

Therefore,

$$z = R^T w = R^T Q^T Q w = A^T (Qw),$$

so

$$z \in \mathcal{C}(A^T).$$

Thus,

$$\mathcal{C}(A^T) = \mathcal{C}(R^T).$$

Since R has r linearly independent rows,

$$\text{rank}(R^T) = r.$$

Therefore,

$$\text{rank}(A^T) = r = \text{rank}(A).$$

□

8 Rank and Nullity

Let

$$r = \text{rank}(A).$$

Since

$$\dim \mathcal{C}(A) = r,$$

and

$$\text{rank}(A) = \text{rank}(A^T),$$

we also have

$$\dim \mathcal{C}(A^T) = r.$$

Using the orthogonal decompositions above gives

$$\dim \mathcal{N}(A) = n - r$$

and

$$\dim \mathcal{N}(A^T) = m - r.$$

Thus,

Subspace	Dimension
$\mathcal{C}(A)$	r
$\mathcal{C}(A^T)$	r
$\mathcal{N}(A)$	$n - r$
$\mathcal{N}(A^T)$	$m - r$

This gives the familiar rank–nullity identity

$$\boxed{\text{rank}(A) + \dim \mathcal{N}(A) = n.}$$

9 Least Squares Revisited

The four fundamental subspaces give another way to understand least squares.

For a least-squares solution $\hat{x}$,

$$r = b - A\hat{x}$$

satisfies

$$r \perp \mathcal{C}(A).$$

Since

$$\mathcal{C}(A)^\perp = \mathcal{N}(A^T),$$

we have

$$\boxed{r \in \mathcal{N}(A^T)}.$$

Equivalently,

$$A^T r = 0.$$

Substituting

$$r = b - A\hat{x}$$

again gives

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

or

$$A^T A \hat{x} = A^T b.$$

Thus the geometry of orthogonal projection, the normal equations, and the four fundamental subspaces all describe the same least-squares problem from different viewpoints.

10 A Brief Review of Complex Numbers

So far, we have worked entirely over $\mathbb{R}$. Most of linear algebra works almost identically over $\mathbb{C}$, with one important modification: the inner product uses complex conjugation.

10.1 Complex Numbers

A complex number has the form

$$z = a + bi, \quad a, b \in \mathbb{R},$$

where

$$i^2 = -1.$$

The complex conjugate of z is

$$\bar{z} = a - bi.$$

Its magnitude is

$$|z| = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}.$$

For example,

$$z = 3 + 4i \quad \implies \quad \bar{z} = 3 - 4i, \quad |z| = 5.$$

10.2 Conjugate Transpose

For a complex vector or matrix, we replace the transpose by the *conjugate transpose*.

If

$$A = (a_{jk}),$$

then

$$A^* = \overline{A}^T.$$

For example,

$$A = \begin{pmatrix} 1 & i \\ 2 & 1 - i \end{pmatrix}$$

has

$$A^* = \begin{pmatrix} 1 & 2 \\ -i & 1 + i \end{pmatrix}.$$

10.3 Inner Products over $\mathbb{C}$

For $x, y \in \mathbb{C}^n$, the standard inner product is

$$\boxed{\langle x, y \rangle = x^* y.}$$

Thus,

$$\langle x, y \rangle = \sum_{j=1}^n \overline{x_j} y_j.$$

The conjugation is essential. In particular,

$$\langle x, x \rangle = x^* x = \sum_{j=1}^n |x_j|^2 \geq 0.$$

Therefore,

$$\|x\|_2 = \sqrt{x^* x}.$$

Let

$$x = \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

Then

$$x^* = (1 \quad -i),$$

so

$$x^* x = 1 + (-i)i = 2.$$

Hence

$$\|x\|_2 = \sqrt{2}.$$

10.4 What Changes from $\mathbb{R}$ to $\mathbb{C}$?

Most formulas remain the same if we replace every transpose T by conjugate transpose $*$.

For example,

$$Q^T Q = I \quad \longrightarrow \quad Q^* Q = I,$$

$$P_L = Q Q^T \quad \longrightarrow \quad P_L = Q Q^*,$$

$$A^T A \hat{x} = A^T b \quad \longrightarrow \quad A^* A \hat{x} = A^* b.$$

A real symmetric matrix satisfies

$$A = A^T.$$

Its complex analogue is a *Hermitian matrix*, which satisfies

$$A = A^*.$$

Thus, the main rule to remember is

$$\boxed{T \longrightarrow *}$$

when moving from real linear algebra to complex linear algebra.