LINEAR ALGEBRA (MT–121)

CHAPTER 4

ORTHOGONALITY

Aftab Alam

19th April 2024 (Friday)

Spring 2024 Semester

Lecture 18

When A has independent columns, $A^{\mathrm{T}}A$ is square, symmetric, and invertible.

To repeat for emphasis: $A^{\mathrm{T}}A$ is (n by m) times (m by n). Then $A^{\mathrm{T}}A$ is square (n by n). It is symmetric, because its transpose is $(A^{\mathrm{T}}A)^{\mathrm{T}} = A^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{T}}$ which equals $A^{\mathrm{T}}A$. We just proved that $A^{\mathrm{T}}A$ is invertible—provided A has independent columns. Watch the difference between dependent and independent columns:

$$\underset{A^{\mathrm{T}}}{\begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 0 \end{bmatrix}} \underset{\text{dependent}}{\overset{A}{\begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 0 \end{bmatrix}}} = \underset{\text{singular}}{\overset{A^{\mathrm{T}}A}{\begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}}} \qquad \underset{A^{\mathrm{T}}}{\begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix}} \underset{\text{indep.}}{\overset{A}{\begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}}} = \underset{\text{invertible}}{\overset{A^{\mathrm{T}}A}{\begin{bmatrix} 2 & 4 \\ 4 & 9 \end{bmatrix}}}$$

Very brief summary To find the projection $\boldsymbol{p} = \widehat{x}_1\boldsymbol{a}_1 + \cdots + \widehat{x}_n\boldsymbol{a}_n$, solve $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. This gives $\widehat{\boldsymbol{x}}$. The projection is $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$ and the error is $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \boldsymbol{b} - A\widehat{\boldsymbol{x}}$. The projection matrix $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ gives $\boldsymbol{p} = P\boldsymbol{b}$.

This matrix satisfies $P^2 = P$. *The distance from b to the subspace $C(A)$ is $||e||$.*

■ REVIEW OF THE KEY IDEAS ■

1. The projection of $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ is $\boldsymbol{p} = \boldsymbol{a}\widehat{\boldsymbol{x}} = \boldsymbol{a}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a})$.

2. The rank one projection matrix $P = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ multiplies $\boldsymbol{b}$ to produce $\boldsymbol{p}$.

3. Projecting $\boldsymbol{b}$ onto a subspace leaves $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ perpendicular to the subspace.

4. When A has full rank n, the equation $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ leads to $\widehat{\boldsymbol{x}}$ and $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$.

5. The projection matrix $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ has $P^{\mathrm{T}} = P$ and $P^2 = P$ and $P\boldsymbol{b} = \boldsymbol{p}$.

WORKED EXAMPLE

4.2 A Project the vector $\boldsymbol{b} = (3,4,4)$ onto the line through $\boldsymbol{a} = (2,2,1)$ and then onto the plane that also contains $\boldsymbol{a}^* = (1,0,0)$. Check that the first error vector $\boldsymbol{b} - \boldsymbol{p}$ is perpendicular to $\boldsymbol{a}$, and the second error vector $\boldsymbol{e}^* = \boldsymbol{b} - \boldsymbol{p}^*$ is also perpendicular to $\boldsymbol{a}^*$.

Find the 3 by 3 projection matrix P onto that plane of $\boldsymbol{a}$ and $\boldsymbol{a}^*$. Find a vector whose *projection onto the plane is the zero vector*. Why is it exactly the error $\boldsymbol{e}^*$?

Solution The projection of $\boldsymbol{b} = (3,4,4)$ onto the line through $\boldsymbol{a} = (2,2,1)$ is $\boldsymbol{p} = 2\boldsymbol{a}$:

Onto a line $$\boldsymbol{p} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a} = \frac{18}{9}(2,2,1) = (4,4,2) = 2\boldsymbol{a}.$$

The error vector $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = (-1,0,2)$ is perpendicular to $\boldsymbol{a} = (2,2,1)$. So $\boldsymbol{p}$ is correct.

The plane of $\boldsymbol{a} = (2,2,1)$ and $\boldsymbol{a}^* = (1,0,0)$ is the column space of $A = [\boldsymbol{a}\ \boldsymbol{a}^*]$:

$$A = \begin{bmatrix} 2 & 1 \\ 2 & 0 \\ 1 & 0 \end{bmatrix} \quad A^{\mathrm{T}}A = \begin{bmatrix} 9 & 2 \\ 2 & 1 \end{bmatrix} \quad (A^{\mathrm{T}}A)^{-1} = \frac{1}{5}\begin{bmatrix} 1 & -2 \\ -2 & 9 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & .8 & .4 \\ 0 & .4 & .2 \end{bmatrix}$$

Now $\boldsymbol{p}^* = P\boldsymbol{b} = (3, 4.8, 2.4)$. The error $\boldsymbol{e}^* = \boldsymbol{b} - \boldsymbol{p}^* = (0, -.8, 1.6)$ is perpendicular to $\boldsymbol{a}$ and $\boldsymbol{a}^*$. This $\boldsymbol{e}^*$ is in the nullspace of P and *its projection is zero*! Note $P^2 = P = P^{\mathrm{T}}$.

Project b onto the column space of A by solving $A^T A\hat{x} = A^T$ b and p = $A\hat{x}$.

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

Find $e = b - p$. It should be perpendicular to the column space of A.

$$A^T A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1+0+0 & 1+0+0 \\ 1+0+0 & 1+1+0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2+0+0 \\ 2+3+0 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$(A^T A)^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

$$x = (A^T A)^{-1} A^T b$$

$$x = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

$$p = Ax = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$$

$$p = A(A^T A)^{-1} A^T b = (2, 3, 0)$$

$$e = b - p = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$$

$$A^T e = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} = 0$$

Project b onto the column space of A by solving $A^T A\hat{x} = A^T$ b and p = $A\hat{x}$.

A = $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$ and b = $\begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$

Find $e = b - p$. It should be perpendicular to the column space of A.

$$A^T A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 8 \\ 14 \end{bmatrix}$$

$$(A^T A)^{-1} = \begin{bmatrix} \frac{3}{2} & -1 \\ -1 & 1 \end{bmatrix}$$

$$x = (A^T A)^{-1} A^T b$$

$$x = \begin{bmatrix} \frac{3}{2} & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 8 \\ 14 \end{bmatrix} = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$$

$$p = Ax = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$$

$$p = A(A^T A)^{-1} A^T b = (4, 4, 6)$$

$$e = b - p = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$A^T e = 0$$

$e = (0, 0, 0)$ means that b is in the column space of A.

Suppose A is the 4 by 4 identity matrix with its last column removed. A is 4 by 3. Project b = (1, 2, 3, 4) onto the column space of A. What shape is the projection matrix P and what is p?

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$P = A(A^T A)^{-1} A^T$$

$$A^T A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

$$P = A\,A^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = I$$

$$p = Pb = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

If A is doubled, then $P = 2A(4A^TA)^{-1}2A^T$. This is the same as P= $A\,(A^TA)^{-1}A^T$.
The column space of 2A is the same as __________ .

Is x the same for A and 2A?

2A has the same column space as A. Then P is same for A and 2A.

$$P_A = P_{2A}$$

$\hat{x}$ for A is $\widehat{x1}$ = A$(A^TA)^{-1}\ A^T$b
$\hat{x}$ for 2A is $\widehat{x2}$ = $(2A^T2A)^{-1}\ 2A^Tb$
$\Rightarrow \widehat{x2}$= $\frac{2}{4}(A^TA)^{-1}\ A^T$b = $\frac{1}{2}\widehat{x1}$
Therefore $\hat{x}$ for 2A is half of $\hat{x}$ for A.

PROJECTIONS: SUMMARY

The key step was $A^{\mathrm{T}}(\boldsymbol{b} - A\widehat{\boldsymbol{x}}) = \mathbf{0}$. We used geometry ($\boldsymbol{e}$ is orthogonal to each $\boldsymbol{a}$). *Linear algebra gives this "normal equation" too*, in a very quick and beautiful way:

1. Our subspace is the column space of A.
2. The error vector $\boldsymbol{b} - A\widehat{\boldsymbol{x}}$ is perpendicular to that column space.
3. Therefore $\boldsymbol{b} - A\widehat{\boldsymbol{x}}$ is in the nullspace of A^{T}! This means $A^{\mathrm{T}}(\boldsymbol{b} - A\widehat{\boldsymbol{x}}) = \mathbf{0}$.

The left nullspace is important in projections. That nullspace of A^{T} contains the error vector $\boldsymbol{e} = \boldsymbol{b} - A\widehat{\boldsymbol{x}}$. The vector $\boldsymbol{b}$ is being split into the projection $\boldsymbol{p}$ and the error $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$. Projection produces a right triangle with sides $\boldsymbol{p}$, $\boldsymbol{e}$, and $\boldsymbol{b}$.

LEAST SQUARE APPROXIMATIONS

1 Solving $\boxed{A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}}$ gives the projection $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$ of $\boldsymbol{b}$ onto the column space of A.

2 When $A\boldsymbol{x} = \boldsymbol{b}$ has no solution, $\widehat{\boldsymbol{x}}$ is the "least-squares solution": $||\boldsymbol{b} - A\widehat{\boldsymbol{x}}||^2 =$ minimum.

3 Setting partial derivatives of $E = ||A\boldsymbol{x} - \boldsymbol{b}||^2$ to zero $\left(\frac{\partial E}{\partial x_i} = 0\right)$ also produces $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

4 To fit points $(t_1, b_1), \ldots, (t_m, b_m)$ by a straight line, A has columns $(1, \ldots, 1)$ and $(t_1, \ldots, t_m)$.

5 In that case $A^{\mathrm{T}}A$ is the 2 by 2 matrix $\begin{bmatrix} m & \boldsymbol{\Sigma}\, t_i \\ \boldsymbol{\Sigma}\, t_i & \boldsymbol{\Sigma}\, t_i^2 \end{bmatrix}$ and $A^{\mathrm{T}}\boldsymbol{b}$ is the vector $\begin{bmatrix} \boldsymbol{\Sigma}\, b_i \\ \boldsymbol{\Sigma}\, t_i b_i \end{bmatrix}$.

LEAST SQUARE APPROXIMATIONS

What if m > n?

- more equations than unknowns.
- n columns span a small part of m-dimensional space.
- b is outside the column space of A.
- Elimination reaches an impossible equation and stops.
- We can't always get the error e = b – A*x* equal to 0.
- When e = 0, x has an exact solution to A*x* = b.
- When the length of e is as small as possible, *x* is a least square solution.

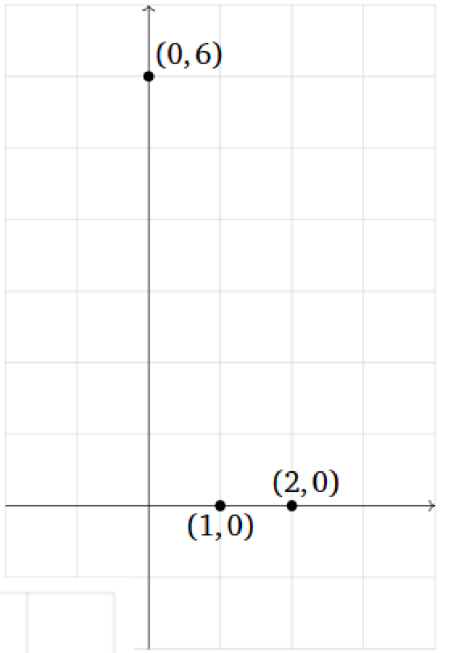
When $Ax = b$ has no solution, multiply by A^{T} and solve $A^{\mathrm{T}}A\hat{x} = A^{\mathrm{T}}b$.

COMPUTING THE LEAST SQUARE SOLUTION

Let A be an $m \times n$ matrix and let b be a vector in R^n. Here is the method for computing a least-squares solution of $Ax = b$:

1. Compute the matrix A^TA and the vector A^Tb.
2. Form the augmented matrix for the matrix equation $A^TAx = A^Tb$, and row reduce.
3. This equation is always consistent, and any solution x is a least-squares solution.

Example 1 A crucial application of least squares is fitting a straight line to m points. Start with three points: *Find the closest line to the points* $(0,6)$, $(1,0)$, *and* $(2,0)$.



No straight line $b = C + Dt$ goes through those three points. We are asking for two numbers C and D that satisfy three equations: $n = 2$ and $m = 3$. Here are the three equations at $t = 0, 1, 2$ to match the given values $b = 6, 0, 0$:

$t = 0$	The first point is on the line $b = C + Dt$ if	$C + D \cdot 0 = 6$
$t = 1$	The second point is on the line $b = C + Dt$ if	$C + D \cdot 1 = 0$
$t = 2$	The third point is on the line $b = C + Dt$ if	$C + D \cdot 2 = 0.$

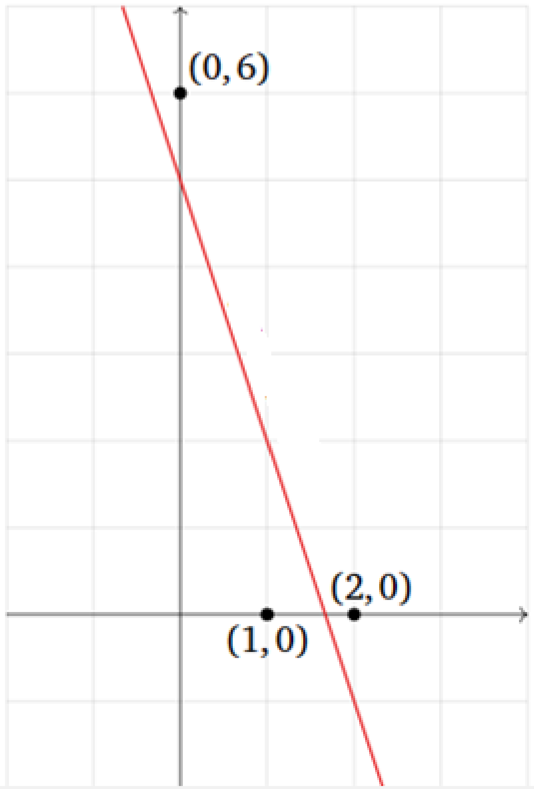
This 3 by 2 system has *no solution*: $\boldsymbol{b} = (6,0,0)$ is not a combination of the column $(1,1,1)$ and $(0,1,2)$. Read off A, $\boldsymbol{x}$, and $\boldsymbol{b}$ from those equations:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} C \\ D \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \quad A\boldsymbol{x} = \boldsymbol{b} \text{ is } \textit{not} \text{ solvable.}$$

$$A^{\mathrm{T}}A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \quad \text{and} \quad A^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

Now solve the normal equation $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$ to find $\widehat{x}$:

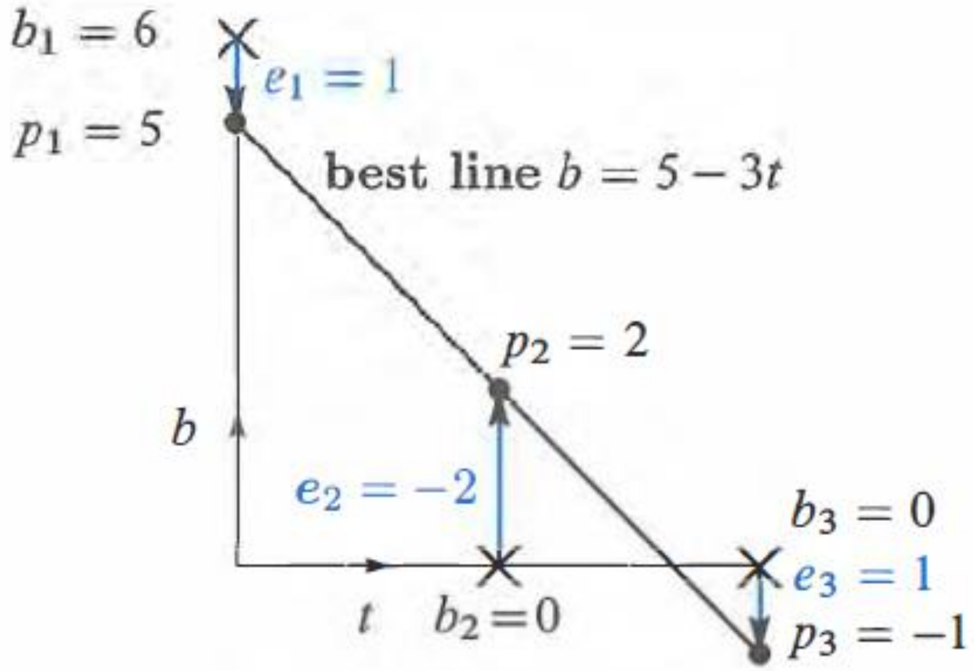
$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \quad \text{gives} \quad \widehat{x} = \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{5} \\ -\mathbf{3} \end{bmatrix}$$



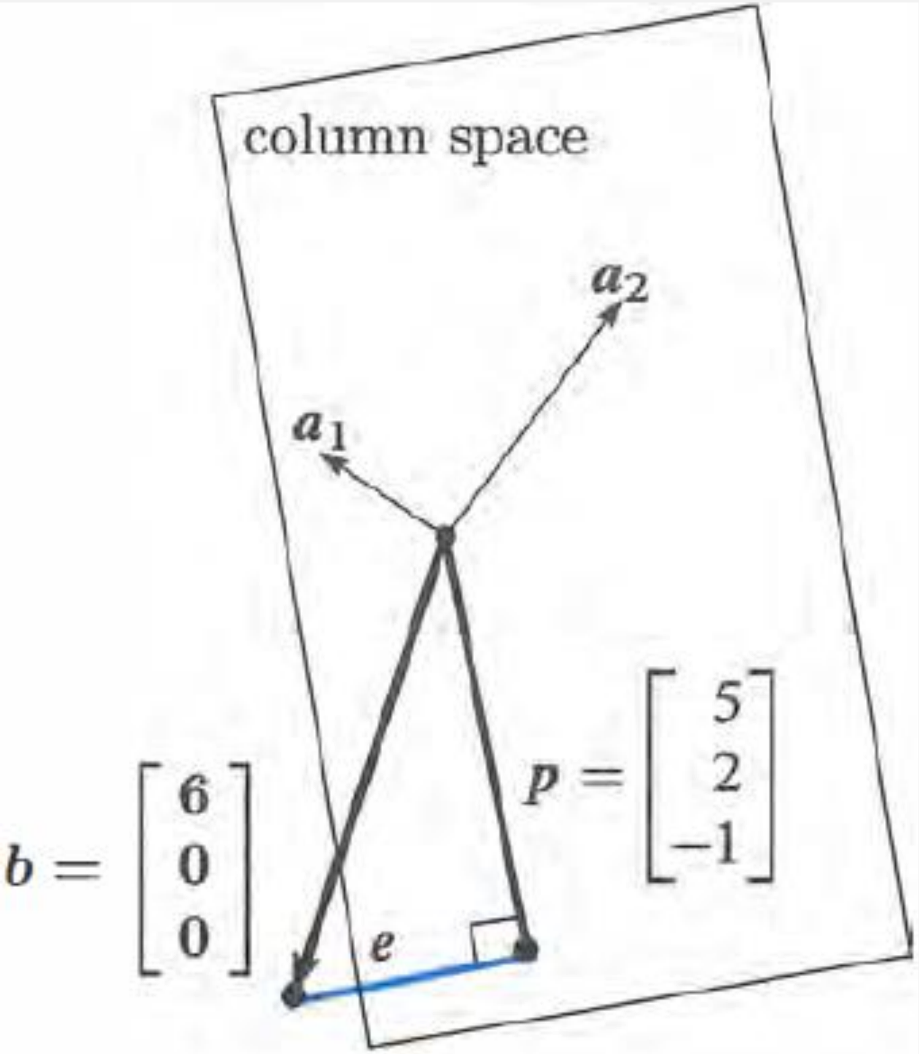
$5 - 3t$ will be the best line for the 3 points.

$$p = A\,\hat{x} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ -3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

$$e = b - p = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$



errors = vertical distances to line



$e = (1, -2, 1)$

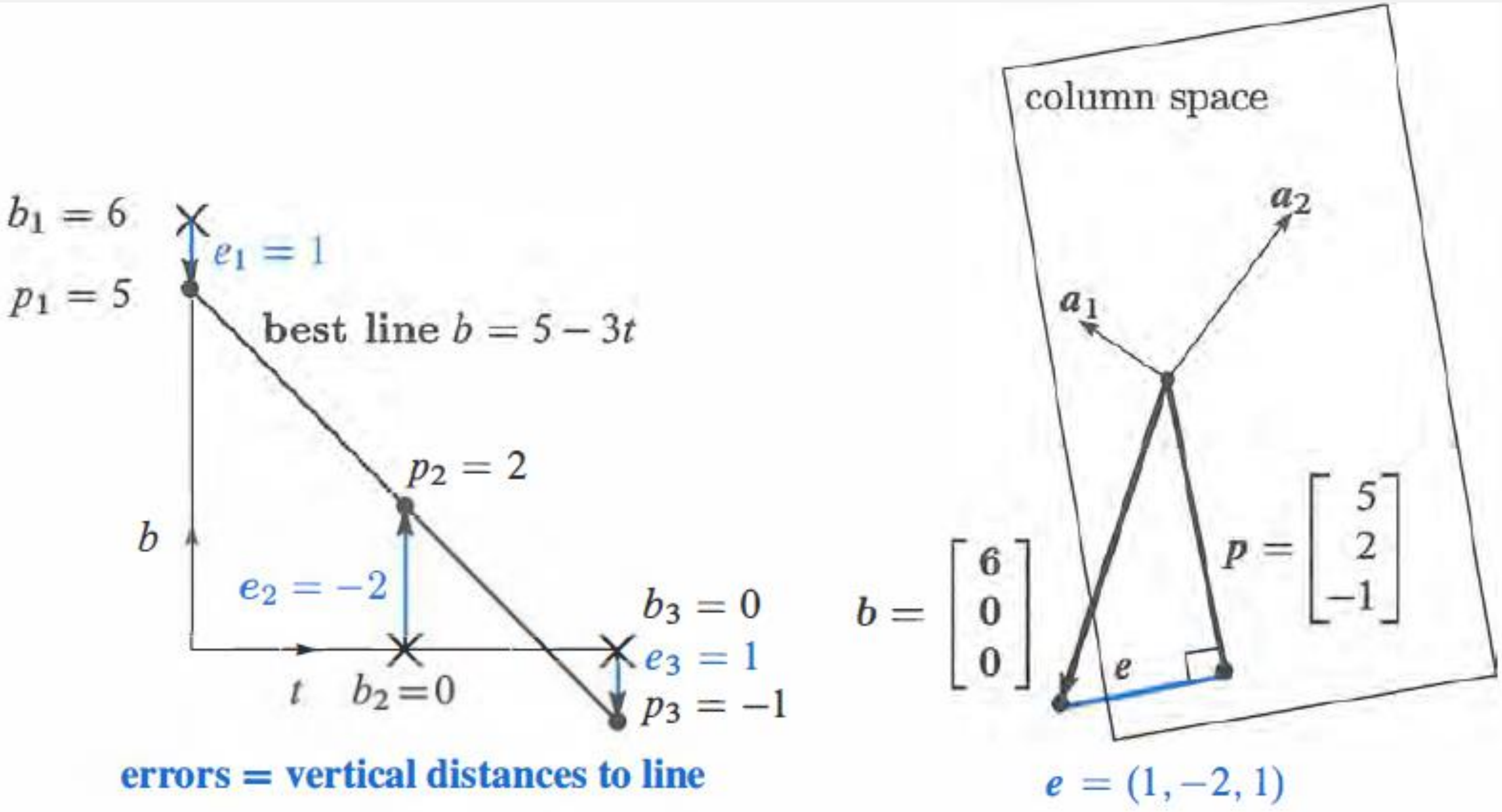


Figure 4.6: **Best line and projection: Two pictures, same problem.** The line has heights $\boldsymbol{p} = (5, 2, -1)$ with errors $\boldsymbol{e} = (1, -2, 1)$. The equations $A^{\mathrm{T}} A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ give $\widehat{\boldsymbol{x}} = (5, -3)$. Same answer! The best line is $b = 5 - 3t$ and the closest point is $\boldsymbol{p} = 5\boldsymbol{a}_1 - 3\boldsymbol{a}_2$.

Notice that the errors $1, -2, 1$ add to zero. *Reason*: The error $\boldsymbol{e} = (e_1, e_2, e_3)$ is perpendicular to the first column $(1, 1, 1)$ in A. The dot product gives $e_1 + e_2 + e_3 = 0$.

Orthonormal Bases and Gram-Schmidt

- **Orthonormal bases** and the **Gram-Schmidt** process are fundamental concepts in LA used to construct orthonormal bases from a set of linearly independent vectors.
- The Gram-Schmidt process is a crucial tool for constructing orthonormal bases in various mathematical contexts, including **solving systems of linear equations**, **finding eigenvalues and eigenvectors**, and **performing orthogonalization in inner product spaces**.
- Gram-Schmidt ensures that a set of linearly independent vectors can be transformed into a set of **orthogonal vectors** and further normalized to form an **orthonormal bases**.

Orthonormal Bases

- An orthonormal bases is a set of vectors in a vector space that are both **orthogonal** (perpendicular to each other) and **normalized** (having unit length).
  - the dot product of any two distinct vectors is 0
  - the norm (length) of each vector is 1.

- Orthonormal bases are particularly useful in various mathematical and engineering applications
  - signal processing
  - quantum mechanics
  - computer graphics

Gram-Schmidt Process

- The Gram-Schmidt process is a method for orthonormalizing a set of linearly independent vectors in a vector space to construct an orthonormal basis.
- Given a set of linearly independent vectors {v1, v2, …, vn}, the Gram-Schmidt process iteratively constructs orthogonal vectors {u1 ,u2 ,…,un }, which are subsequently normalized to form an orthonormal basis.
- The process involves subtracting from each vector its projection onto the subspace spanned by the previously orthonormalized vectors.

Let $\mathbf{u}_1 = \mathbf{v}_1$.

For $i = 2, 3, \ldots, n$:

- $\mathbf{u}_i = \mathbf{v}_i - \sum_{j=1}^{i-1} \mathrm{proj}_{\mathbf{u}_j}(\mathbf{v}_i)$, where $\mathrm{proj}_{\mathbf{u}_j}(\mathbf{v}_i)$ is the projection of $\mathbf{v}_i$ onto $\mathbf{u}_j$.

Normalize each $\mathbf{u}_i$ to obtain the orthonormal basis vectors.

Orthonormal Bases and Gram-Schmidt

1 The columns $q_1, \ldots, q_n$ are orthonormal if $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j = \left\{ \begin{array}{l} 0 \text{ for } i \neq j \\ 1 \text{ for } i = j \end{array} \right\}$. Then $\boxed{Q^{\mathrm{T}}Q = I\ .}$

2 If Q is also square, then $QQ^{\mathrm{T}} = I$ and $\boxed{Q^{\mathrm{T}} = Q^{-1}}$. Q is an "orthogonal matrix".

3 The least squares solution to $Q\boldsymbol{x} = \boldsymbol{b}$ is $\widehat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$. Projection of $\boldsymbol{b}$: $\boldsymbol{p} = QQ^{\mathrm{T}}\boldsymbol{b} = P\boldsymbol{b}$.

4 The **Gram-Schmidt** process takes independent $\boldsymbol{a}_i$ to orthonormal $\boldsymbol{q}_i$. Start with $\boldsymbol{q}_1 = \boldsymbol{a}_1 / ||\boldsymbol{a}_1||$.

5 $\boldsymbol{q}_i$ is $(\boldsymbol{a}_i -$ projection $\boldsymbol{p}_i) \,/\, ||\boldsymbol{a}_i - \boldsymbol{p}_i||$; projection $\boldsymbol{p}_i = (\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{q}_1)\boldsymbol{q}_1 + \cdots + (\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{q}_{i-1})\boldsymbol{q}_{i-1}$.

6 Each $\boldsymbol{a}_i$ will be a combination of $\boldsymbol{q}_1$ to $\boldsymbol{q}_i$. Then $\boldsymbol{A} = \boldsymbol{QR}$: orthogonal Q and triangular R.

Orthonormal Bases

A matrix Q with orthonormal columns satisfies $\boxed{Q^{\mathrm{T}}Q = I}$:

$$\boxed{Q^{\mathrm{T}}Q} = \begin{bmatrix} - & q_1^{\mathrm{T}} & - \\ - & q_2^{\mathrm{T}} & - \\ & & \\ - & q_n^{\mathrm{T}} & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ q_1 & q_2 & & q_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \boxed{I}.$$

Example 1 **(Rotation)** Q rotates every vector in the plane by the angle θ:

$$\boxed{Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}} \quad \text{and} \quad \boxed{Q^{\mathrm{T}} = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}}.$$

The columns of Q are orthogonal (take their dot product). They are unit vectors because $\sin^2\theta + \cos^2\theta = 1$. Those columns give an ***orthonormal basis*** for the plane $\mathbf{R}^2$.

The standard basis vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ are rotated through θ (see Figure 4.10a). Q^{-1} rotates vectors back through $-\theta$. It agrees with Q^{T}, because the cosine of $-\theta$ equals the cosine of θ, and $\sin(-\theta) = -\sin\theta$. We have $Q^{\mathrm{T}}Q = I$ and $QQ^{\mathrm{T}} = I$.

Example 2 **(Permutation)** These matrices change the order to (y, z, x) and (y, x):

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

All columns of these Q's are unit vectors (their lengths are obviously 1). They are also orthogonal (the 1's appear in different places). *The inverse of a permutation matrix is its transpose*: $Q^{-1} = Q^{\mathrm{T}}$. The inverse puts the components back into their original order:

$$\textbf{Inverse = transpose:} \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Every permutation matrix is an orthogonal matrix.

Example 4 The columns of this orthogonal Q are orthonormal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$:

$$\boldsymbol{m = n = 3} \qquad Q = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix} \quad \text{has} \quad Q^{\mathrm{T}}Q = QQ^{\mathrm{T}} = I.$$

The separate projections of $\boldsymbol{b} = (0, 0, 1)$ onto $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ and $\boldsymbol{q}_3$ are $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$:

$$\boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) = \tfrac{2}{3}\boldsymbol{q}_1 \quad \text{and} \quad \boldsymbol{q}_2(\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b}) = \tfrac{2}{3}\boldsymbol{q}_2 \quad \text{and} \quad \boldsymbol{q}_3(\boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{b}) = -\tfrac{1}{3}\boldsymbol{q}_3.$$

The sum of the first two is the projection of $\boldsymbol{b}$ onto the *plane* of $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$. The sum of all three is the projection of $\boldsymbol{b}$ onto the *whole space*—which is $\boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_3 = \boldsymbol{b}$ itself:

Reconstruct $\boldsymbol{b}$
$\boldsymbol{b} = \boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_3$

$$\frac{2}{3}\boldsymbol{q}_1 + \frac{2}{3}\boldsymbol{q}_2 - \frac{1}{3}\boldsymbol{q}_3 = \frac{1}{9}\begin{bmatrix} -2+4-2 \\ 4-2-2 \\ 4+4+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{b}.$$