THEOREM 1

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then

a. $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$

b. $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{w} + \mathbf{v}\cdot\mathbf{w}$

c. $(c\mathbf{u})\cdot\mathbf{v} = c(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot(c\mathbf{v})$

d. $\mathbf{u}\cdot\mathbf{u} \geq 0$, and $\mathbf{u}\cdot\mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Properties (b) and (c) can be combined several times to produce the following useful rule:

$$(c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{w} = c_1(\mathbf{u}_1\cdot\mathbf{w}) + \cdots + c_p(\mathbf{u}_p\cdot\mathbf{w})$$

DEFINITION

The **length** (or **norm**) of $\mathbf{v}$ is the nonnegative scalar $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \quad \text{and} \quad \|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v}$$

DEFINITION

For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the **distance between u and v**, written as $\text{dist}(\mathbf{u},\mathbf{v})$, is the length of the vector $\mathbf{u}-\mathbf{v}$. That is,

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

THEOREM 2

The Pythagorean Theorem

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

$+/-\ 2 V\cdot U$

$\downarrow$

$=0$

$\|U+V\|^2 = (U+V)\cdot(U+V)$

$= U\cdot(U+V) + V\cdot(U+V)$

$= V\cdot U + U\cdot V + V\cdot U + V\cdot V$

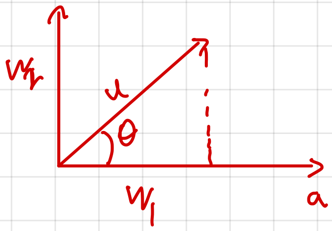
$= 2V\cdot U + \|U\|^2 + \|V\|^2$ if $\perp$, $V\cdot U = 0$

THEOREM 3

Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

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

Projection Theorem



Proj of u along a

$\text{Proj}_a u = \frac{u \cdot a}{\|a\|^2} a = \hat{u}$

$W_2 = u - \text{proj}_a u$

$W_2 \cdot W_1 = 0$

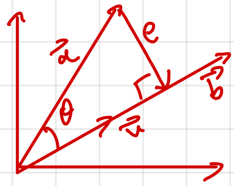
$\|v\| = \sqrt{v^T v}$

$\text{Proj}_v v = v\left(\frac{v^T v}{\|v\|^2}\right)$

$\|u+v\|^2 = \underbrace{\|u\|^2 + \|v\|^2}$

Proof: $(u+v)(u+v) = \|u\|^2 + 2(\overset{=0}{u \cdot v}) + \|v\|^2$

Proof (method 1)



$\cos\theta = \frac{|\vec{u}|}{|\vec{a}|}$ $\cos\theta = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}$

$\frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|} = \frac{|\vec{u}|}{|\vec{a}|}$

$\boxed{\vec{u} = |\vec{u}|\hat{u}}$

$|\vec{u}| = \frac{\vec{a}\cdot\vec{b}}{\cancel{|\vec{a}|}|\vec{b}|}\cancel{|\vec{a}|} = \frac{\vec{a}\cdot\vec{b}}{|\vec{b}|}$

$\vec{u} = \frac{a \cdot b}{|\vec{b}|}\boxed{\frac{1}{|\vec{b}|}\vec{b}} = \frac{\vec{a}\cdot\vec{b}}{|b|^2}\vec{b} = \text{proj}_b \vec{a}$

$\hat{u} = \hat{b}$

Method 2

$e = a - u$

e is orthogonal to b

$e \cdot b = 0$

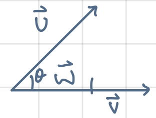
$(a-u)\cdot b = 0$

$a \cdot b = u \cdot b$

$(a \cdot b) b = (u \cdot b) b$

$(a \cdot b) b = u (b \cdot b)$

$u = \frac{(a \cdot b) b}{b \cdot b}$



cah

$\cos\theta = \frac{\|\vec{w}\|}{\|\vec{u}\|}$

$\|\vec{w}\| = \|\vec{u}\| \cos\theta$

$u \cdot v = \|\vec{u}\|\|\vec{v}\| \cos\theta$

$\|\vec{w}\| = \frac{\|\vec{u}\|\|\vec{v}\| \cos\theta}{\|\vec{v}\|}$

$\|\vec{w}\| = \frac{u \cdot v}{\|\vec{v}\|}$

$\vec{w} = \frac{u \cdot v}{\|\vec{v}\|} \cdot \frac{\vec{v}}{\|\vec{v}\|}$

$= \left(\frac{u \cdot v}{\|\vec{v}\|^2}\right) \cdot \vec{v}$

$= \frac{u \cdot v}{v \cdot v} \cdot v$

THEOREM 4 *Note: $p \le n$*

If $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

PROOF If $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1, \dots, c_p$, then

$$\begin{aligned} 0 = \mathbf{0}\cdot\mathbf{u}_1 &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{u}_1 \\ &= (c_1\mathbf{u}_1)\cdot\mathbf{u}_1 + (c_2\mathbf{u}_2)\cdot\mathbf{u}_1 + \cdots + (c_p\mathbf{u}_p)\cdot\mathbf{u}_1 \\ &= c_1(\mathbf{u}_1\cdot\mathbf{u}_1) + c_2(\mathbf{u}_2\cdot\mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p\cdot\mathbf{u}_1) \\ &= c_1(\mathbf{u}_1\cdot\mathbf{u}_1) \end{aligned}$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \dots, \mathbf{u}_p$. Since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1\cdot\mathbf{u}_1$ is not zero and so $c_1 = 0$. Similarly, $c_2, \dots, c_p$ must be zero. Thus S is linearly independent. ■

THEOREM 5

Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$$

are given by

$$c_j = \frac{\mathbf{y}\cdot\mathbf{u}_j}{\mathbf{u}_j\cdot\mathbf{u}_j} \qquad (j = 1, \dots, p)$$

PROOF As in the preceding proof, the orthogonality of $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ shows that

$$\mathbf{y}\cdot\mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{u}_1 = c_1(\mathbf{u}_1\cdot\mathbf{u}_1)$$

Since $\mathbf{u}_1\cdot\mathbf{u}_1$ is not zero, the equation above can be solved for c_1. To find c_j for $j = 2, \dots, p$, compute $\mathbf{y}\cdot\mathbf{u}_j$ and solve for c_j. ■

THEOREM 6

An $m \times n$ matrix U has orthonormal columns if and only if $U^TU = I$.

THEOREM 7

Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$. Then

a. $\|U\mathbf{x}\| = \|\mathbf{x}\|$

b. $(U\mathbf{x})\cdot(U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$

c. $(U\mathbf{x})\cdot(U\mathbf{y}) = 0$ if and only if $\mathbf{x}\cdot\mathbf{y} = 0$

Orthogonal Matrix

In linear algebra, an **orthogonal matrix**, or **orthonormal matrix**, is a real square matrix whose columns and rows are orthonormal vectors.

***

One way to express this is

$$Q^{\mathrm{T}}Q = QQ^{\mathrm{T}} = I,$$

where Q^{T} is the transpose of Q and I is the identity matrix.

This leads to the equivalent characterization: a matrix Q is orthogonal if its transpose is equal to its inverse:

$$Q^{\mathrm{T}} = Q^{-1},$$

where Q^{-1} is the inverse of Q.

THEOREM 8

The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^{\perp}$. In fact, if $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p \tag{2}$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

THEOREM 9

The Best Approximation Theorem

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

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \tag{3}$$

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

If $\mathbf{y}$ is in $W = \text{Span}\,\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$, then $\text{proj}_W\,\mathbf{y} = \mathbf{y}$.

THEOREM 10

If $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W\,\mathbf{y} = (\mathbf{y}\cdot\mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y}\cdot\mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y}\cdot\mathbf{u}_p)\mathbf{u}_p \tag{4}$$

If $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p\,]$, then

$$\text{proj}_W\,\mathbf{y} = UU^T\mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \tag{5}$$

$Note: \; p < n$

Warning: Theorem 12 is for A Having independent column.

THEOREM 12

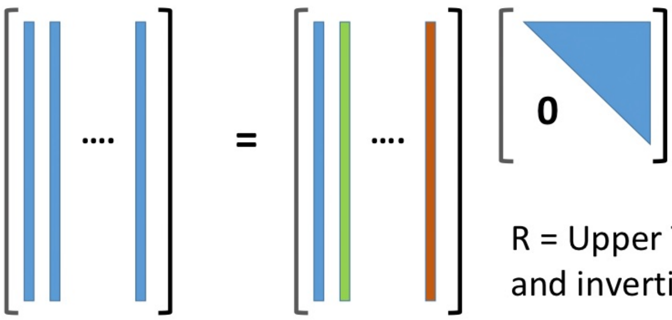
The QR Factorization → Convert normal basis to orthogonal

If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

$$Ax = y$$
$$QRx = y$$
$$Q^TQRx = Q^Ty$$

$$Rx = Q^Ty$$

$$A = [\,\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n\,] = [\,Q\mathbf{r}_1 \quad \cdots \quad Q\mathbf{r}_n\,] = QR$$



R = Upper Triangle and invertible

A $\in R^{m*n}$ Q = columns are orthonormal

- Properties of Q:
  - C(Q) == C(A)
  - $Q^TQ = I$, i.e Q has orthonormal column (BUT not necessary square)
  - QQ^T = projection matrix into col (A)

- R is a square upper triangle matrix and Depending if
  - A has independent col, then R is invertible,
  - A has dependent col, then R is NOT-invertible.

The Gram–Schmidt Process

Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2$$

$$\vdots$$

$$\mathbf{v}_p = \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}$$

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \quad \text{for } 1 \le k \le p \tag{1}$$

Consider A has independent col (mxn matrix)

$$A = [x_1, x_2, x_3, \dots, x_n]$$

Gram–Schmidt Process

$$\text{Proj}_v x = \frac{v.x}{v.v}\, v$$

$$v_1 = x_1$$

$$v_2 = x_2 - \text{Proj}_{v_1} x_2$$

$$v_3 = x_3 - \text{Proj}_{v_1} x_3 - \text{Proj}_{v_2} x_3$$

$$v_i = x_i - \sum_{j=1}^{i-1} \text{Proj}_{v_j} x_i$$

Ortho-normalization

$$u_1 = \frac{v_1}{||v_1||}$$

$$u_2 = \frac{v_2}{||v_2||}$$

$$u_i = \frac{v_i}{||v_i||}$$

We can now express the x_i over our newly computed orthonormal basis: e_j

$$x_1 = u_1.x_1\ u_1$$

$$x_2 = u_1.x_2\ u_1 + u_2.x_2\ u_2$$

$$x_3 = u_1.x_3\ u_1 + u_2.x_3\ u_2 + u_3.x_3\ u_3$$

$$\vdots$$

$$x_n = \sum_{j=1}^{n} u_j.x_n\ u_j$$

This can be written in matrix form:

$$A = QR$$

where:

$$Q = [u_1, u_2, u_3, \dots, u_n]$$

and

$$R = \begin{pmatrix} u_1.x_1 & u_1.x_2 & u_1.x_3 & \cdots \\ 0 & u_2.x_2 & u_2.x_3 & \cdots \\ 0 & 0 & u_3.x_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

$R = Q^T A$

* Don't use this. Use

THEOREM 14

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

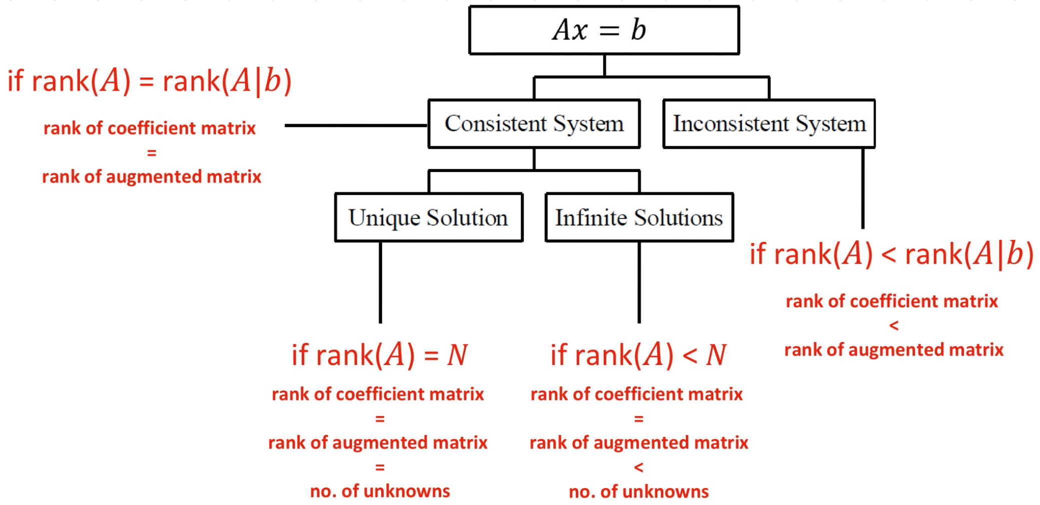
a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly indpendent.

c. The matrix A^TA is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$



THEOREM 6.4.1 Best Approximation Theorem

*If W is a finite-dimensional subspace of an inner product space V, and if $\mathbf{b}$ is a vector in V, then $\text{proj}_W \mathbf{b}$ is the **best approximation** to $\mathbf{b}$ from W in the sense that*

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|$$

for every vector $\mathbf{w}$ in W that is different from $\text{proj}_W \mathbf{b}$.

Proof For every vector $\mathbf{w}$ in W, we can write

$$\mathbf{b} - \mathbf{w} = (\mathbf{b} - \text{proj}_W \mathbf{b}) + (\text{proj}_W \mathbf{b} - \mathbf{w})$$

But $\text{proj}_W \mathbf{b} - \mathbf{w}$, being a difference of vectors in W, is itself in W; and since $\mathbf{b} - \text{proj}_W \mathbf{b}$ is orthogonal to W, the two terms on the right side of (1) are orthogonal. Thus, it follows from the Theorem of Pythagoras (Theorem 6.2.3) that

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathbf{b} - \text{proj}_W \mathbf{b}\|^2 + \|\text{proj}_W \mathbf{b} - \mathbf{w}\|^2$$

If $\mathbf{w} \neq \text{proj}_W \mathbf{b}$, it follows that the second term in this sum is positive, and hence that

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\|^2 < \|\mathbf{b} - \mathbf{w}\|^2$$

Since norms are nonnegative, it follows (from a property of inequalities) that

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\| \quad \blacktriangleleft$$

$Ax = b$

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

Suppose $\hat{\mathbf{x}}$ satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. By the Orthogonal Decomposition Theorem in Section 6.3, the projection $\hat{\mathbf{b}}$ has the property that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col A, so $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of A. If $\mathbf{a}_j$ is any column of A, then $\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0$, and $\mathbf{a}_j^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$. Since each $\mathbf{a}_j^T$ is a row of A^T,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \tag{2}$$

(This equation also follows from Theorem 3 in Section 6.1.) Thus

$$A^T\mathbf{b} - A^TA\hat{\mathbf{x}} = \mathbf{0}$$
$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$$

These calculations show that each least-squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies the equation

$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{3}$$

The matrix equation (3) represents a system of equations called the **normal equations** for $A\mathbf{x} = \mathbf{b}$. A solution of (3) is often denoted by $\hat{\mathbf{x}}$.

Least Squares Solution: → linear regression

From

$$X\beta = y$$

Then pre-multiply by X^T to get the normal equation,

$$X^T X\beta = X^T y$$

Then premultiply by $(X^T X)^{-1}$,

$$(X^T X)^{-1}(X^T X)\beta = \left(X^T X\right)^{-1} X^T y$$
$$\beta = (X^T X)^{-1} X^T y$$

EXAMPLE 2 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$$

singular

SOLUTION Compute

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

The augmented matrix for $A^TA\mathbf{x} = A^T\mathbf{b}$ is

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\hat{b} = A \underbrace{\begin{bmatrix} 3 & -\alpha \\ -5 & \alpha \\ -2 & \alpha \\ 0 & \alpha \end{bmatrix}}_{\hat{x}}$$

The general solution is $x_1 = 3 - x_4$, $x_2 = -5 + x_4$, $x_3 = -2 + x_4$, and x_4 is free. So the general least-squares solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \blacksquare$$

The next theorem gives useful criteria for determining when there is only one least-squares solution of $A\mathbf{x} = \mathbf{b}$. (Of course, the orthogonal projection $\hat{\mathbf{b}}$ is always unique.)