Chapter 4

SECTION 1

2. $x_1 = r\cos\theta$, $x_2 = r\sin\theta$ where $r = (x_1^2 + x_2^2)^{1/2}$ and θ is the angle between $\mathbf{x}$ and $\mathbf{e}_1$.

$$\begin{aligned} L(\mathbf{x}) &= (r\cos\theta\cos\alpha - r\sin\theta\sin\alpha, r\cos\theta\sin\alpha + r\sin\theta\cos\alpha)^T \\ &= (r\cos(\theta+\alpha), r\sin(\theta+\alpha))^T \end{aligned}$$

The linear transformation L has the effect of rotating a vector by an α in the counterclockwise direction.

3. If $\alpha \neq 1$ then

$$L(\alpha\mathbf{x}) = \alpha\mathbf{x} + \mathbf{a} \neq \alpha\mathbf{x} + \alpha\mathbf{a} = \alpha L(\mathbf{x})$$

The addition property also fails

$$\begin{aligned} L(\mathbf{x}+\mathbf{y}) &= \mathbf{x}+\mathbf{y}+\mathbf{a} \\ L(\mathbf{x}) + L(\mathbf{y}) &= \mathbf{x}+\mathbf{y}+2\mathbf{a} \end{aligned}$$

4. Let

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$$

To determine $L(\mathbf{x})$ we must first express $\mathbf{x}$ as a linear combination

$$\mathbf{x} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2$$

To do this we must solve the system $U\mathbf{c} = \mathbf{x}$ for $\mathbf{c}$. The solution is $\mathbf{c} = (4, 3)^T$ and it follows that

$$L(\mathbf{x}) = L(4\mathbf{u}_1 + 3\mathbf{u}_2) = 4L(\mathbf{u}_1) + 3L(\mathbf{u}_2) = 4\begin{pmatrix} -2 \\ 3 \end{pmatrix} + 3\begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} 7 \\ 18 \end{pmatrix}$$

8. **(a)**

$$L(\alpha A) = C(\alpha A) + (\alpha A)C = \alpha(CA + AC) = \alpha L(A)$$

and

$$\begin{aligned} L(A+B) &= C(A+B) + (A+B)C = CA + CB + AC + BC \\ &= (CA + AC) + (CB + BC) = L(A) + L(B) \end{aligned}$$

Therefore L is a linear operator.

(b) $L(\alpha A + \beta B) = C^2(\alpha A + \beta B) = \alpha C^2 A + \beta C^2 B = \alpha L(A) + \beta L(B)$
Therefore L is a linear operator.

(c) If $C \neq O$ then L is not a linear operator. For example,

$$L(2I) = (2I)^2 C = 4C \neq 2C = 2L(I)$$

10. If $f, g \in C[0, 1]$ then

$$\begin{aligned} L(\alpha f + \beta g) &= \int_0^x (\alpha f(t) + \beta g(t))dt \\ &= \alpha \int_0^x f(t)dt + \beta \int_0^x g(t)dt \\ &= \alpha L(f) + \beta L(g) \end{aligned}$$

Thus L is a linear transformation from $C[0, 1]$ to $C[0, 1]$.

12. If L is a linear operator from V into W use mathematical induction to prove

$$L(\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n) = \alpha_1 L(\mathbf{v}_1) + \alpha_2 L(\mathbf{v}_2) + \cdots + \alpha_n L(\mathbf{v}_n).$$

Proof: In the case $n = 1$

$$L(\alpha_1 \mathbf{v}_1) = \alpha_1 L(\mathbf{v}_1)$$

Let us assume the result is true for any linear combination of k vectors and apply L to a linear combination of $k + 1$ vectors.

$$\begin{aligned} L(\alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k + \alpha_{k+1} \mathbf{v}_{k+1}) &= L([\alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k] + [\alpha_{k+1} \mathbf{v}_{k+1}]) \\ &= L(\alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k) + L(\alpha_{k+1} \mathbf{v}_{k+1}) \\ &= \alpha_1 L(\mathbf{v}_1) + \cdots + \alpha_k L(\mathbf{v}_k) + \alpha_{k+1} L(\mathbf{v}_{k+1}) \end{aligned}$$

The result follows then by mathematical induction.

13. If $\mathbf{v}$ is any element of V then

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n$$

Since $L_1(\mathbf{v}_i) = L_2(\mathbf{v}_i)$ for $i = 1, \ldots, n$, it follows that

$$\begin{aligned} L_1(\mathbf{v}) &= \alpha_1 L_1(\mathbf{v}_1) + \alpha_2 L_1(\mathbf{v}_2) + \cdots + \alpha_n L_1(\mathbf{v}_n) \\ &= \alpha_1 L_2(\mathbf{v}_1) + \alpha_2 L_2(\mathbf{v}_2) + \cdots + \alpha_n L_2(\mathbf{v}_n) \\ &= L_2(\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n) \\ &= L_2(\mathbf{v}) \end{aligned}$$

14. Let L be a linear transformation from R^1 to R^1. If $L(\mathbf{1}) = \mathbf{a}$ then

$$L(\mathbf{x}) = L(x\mathbf{1}) = xL(\mathbf{1}) = x\mathbf{a} = a\mathbf{x}$$

15. The proof is by induction on n. In the case $n = 1$, L^1 is a linear operator since $L^1 = L$. We will show that if L^m is a linear operator on V then L^{m+1} is also a linear operator on V. This follows since

$$L^{m+1}(\alpha\mathbf{v}) = L(L^m(\alpha\mathbf{v})) = L(\alpha L^m(\mathbf{v})) = \alpha L(L^m(\mathbf{v})) = \alpha L^{m+1}(\mathbf{v})$$

and

$$\begin{aligned} L^{m+1}(\mathbf{v}_1 + \mathbf{v}_2) &= L(L^m(\mathbf{v}_1 + \mathbf{v}_2)) \\ &= L(L^m(\mathbf{v}_1) + L^m(\mathbf{v}_2)) \\ &= L(L^m(\mathbf{v}_1)) + L(L^m(\mathbf{v}_2)) \\ &= L^{m+1}(\mathbf{v}_1) + L^{m+1}(\mathbf{v}_2) \end{aligned}$$

16. If $\mathbf{v}_1, \mathbf{v}_2 \in V$, then

$$\begin{aligned} L(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) &= L_2(L_1(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2)) \\ &= L_2(\alpha L_1(\mathbf{v}) + \beta L_1(\mathbf{v}_2)) \\ &= \alpha L_2(L_1(\mathbf{v}_1)) + \beta L_2(L_1(\mathbf{v}_2)) \\ &= \alpha L(\mathbf{v}_1) + \beta L(\mathbf{v}_2) \end{aligned}$$

Therefore L is a linear transformation.

17. (b) $\ker(L) = \text{Span}(\mathbf{e}_3)$, $L(R^3) = \text{Span}(\mathbf{e}_1, \mathbf{e}_2)$

18. (c) $L(S) = \text{Span}((1, 1, 1)^T)$

19. (b) If $p(x) = ax^2 + bx + c$ is in $\ker(L)$, then

$$L(p) = (ax^2 + bx + c) - (2ax + b) = ax^2 + (b - 2a)x + (c - b)$$

must equal the zero polynomial $z(x) = 0x^2 + 0x + 0$. Equating coefficients we see that $a = b = c = 0$ and hence $\ker(L) = \{\mathbf{0}\}$. The range of L is all of P_3. To see this note that if $p(x) = ax^2 + bx + c$ is any vector in P_3 and we define $q(x) = ax^2 + (b + 2a)x + c + b + 2a$ then

$$L(q(x)) = (ax^2 + (b+2a)x + c + b + 2a) - (2ax + b + 2a) = ax^2 + bx + c = p(x)$$

20. If $\mathbf{0}_V$ denotes the zero vector in V and $\mathbf{0}_W$ is the zero vector in W then $L(\mathbf{0}_V) = \mathbf{0}_W$. Since $\mathbf{0}_W$ is in T, it follows that $\mathbf{0}_V$ is in $L^{-1}(T)$ and hence $L^{-1}(T)$ is nonempty. If $\mathbf{v}$ is in $L^{-1}(T)$, then $L(\mathbf{v}) \in T$. It follows that $L(\alpha\mathbf{v}) = \alpha L(\mathbf{v})$ is in T and hence $\alpha\mathbf{v} \in L^{-1}(T)$. If $\mathbf{v}_1, \mathbf{v}_2 \in L^{-1}(T)$, then $L(\mathbf{v}_1)$, $L(\mathbf{v}_2)$ are in T and hence

$$L(\mathbf{v}_1 + \mathbf{v}_2) = L(\mathbf{v}_1) + L(\mathbf{v}_2)$$

is also an element of $L(T)$. Thus $\mathbf{v}_1 + \mathbf{v}_2 \in L^{-1}(T)$ and therefore $L^{-1}(T)$ is a subspace of V.

21. Suppose L is one-to-one and $\mathbf{v} \in \ker(L)$.

$$L(\mathbf{v}) = \mathbf{0}_W \qquad \text{and} \qquad L(\mathbf{0}_V) = \mathbf{0}_W$$

Since L is one-to-one, it follows that $\mathbf{v} = \mathbf{0}_V$. Therefore $\ker(L) = \{\mathbf{0}_V\}$.

Conversely, suppose $\ker(L) = \{\mathbf{0}_V\}$ and $L(\mathbf{v}_1) = L(\mathbf{v}_2)$. Then

$$L(\mathbf{v}_1 - \mathbf{v}_2) = L(\mathbf{v}_1) - L(\mathbf{v}_2) = \mathbf{0}_W$$

Therefore $\mathbf{v}_1 - \mathbf{v}_2 \in \ker(L)$ and hence

$$\begin{aligned} \mathbf{v}_1 - \mathbf{v}_2 &= \mathbf{0}_V \\ \mathbf{v}_1 &= \mathbf{v}_2 \end{aligned}$$

So L is one-to-one.

22. To show that L maps R^3 onto R^3 we must show that for any vector $\mathbf{y} \in R^3$ there exists a vector $\mathbf{x} \in R^3$ such that $L(\mathbf{x}) = \mathbf{y}$. This is equivalent to showing that the linear system

$$\begin{array}{lllllll} x_1 & & & & & = & y_1 \\ x_1 & + & x_2 & & & = & y_2 \\ x_1 & + & x_2 & + & x_3 & = & y_3 \end{array}$$

is consistent. This system is consistent since the coefficient matrix is nonsingular.

24. (a) $L(R^2) = \{A\mathbf{x} \mid \mathbf{x} \in R^2\}$

$$\begin{aligned} &= \{x_1\mathbf{a}_1 + x_2\mathbf{a}_2 \mid x_1, x_2 \text{ real }\} \\ &= \text{the column space of } A \end{aligned}$$

(b) If A is nonsingular, then A has rank 2 and it follows that its column space must be R^2. By part (a), $L(R^2) = R^2$.

25. (a) If $p = ax^2 + bx + c \in P_3$, then

$$D(p) = 2ax + b$$

Thus

$$D(P_3) = \text{Span}(1, x) = P_2$$

The operator is not one-to-one, for if $p_1(x) = ax^2 + bx + c_1$ and $p_2(x) = ax^2 + bx + c_2$ where $c_2 \neq c_1$, then $D(p_1) = D(p_2)$.

(b) The subspace S consists of all polynomials of the form $ax^2 + bx$. If $p_1 = a_1x^2 + b_1x$, $p_2 = a_2x^2 + b_2x$ and $D(p_1) = D(p_2)$, then

$$2a_1x + b_1 = 2a_2x + b_2$$

and it follows that $a_1 = a_2$, $b_1 = b_2$. Thus $p_1 = p_2$ and hence D is one-to-one. D does not map S onto P_3 since $D(S) = P_2$.

SECTION 2

7. (a) $\mathcal{I}(\mathbf{e}_1) = 0\mathbf{y}_1 + 0\mathbf{y}_2 + 1\mathbf{y}_3$
$\mathcal{I}(\mathbf{e}_2) = 0\mathbf{y}_1 + 1\mathbf{y}_2 - 1\mathbf{y}_3$
$\mathcal{I}(\mathbf{e}_3) = 1\mathbf{y}_1 - 1\mathbf{y}_2 + 0\mathbf{y}_3$

10. (c) $\begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix}$

11. (a) $YP = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$

(b) $PY = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

(c) $PR = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$

(d) $RP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & -1 & 0 \end{pmatrix}$

(e)

$$\begin{aligned} YPR &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \end{aligned}$$

(f)

$$\begin{aligned} RPY &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \end{aligned}$$

12. (a) If Y is the yaw matrix and we expand $\det(Y)$ along its third row we get

$$\det(Y) = \cos^2 u + \sin^2 u = 1$$

Similarly, if we expand the determinant pitch matrix P along its second and expand the determinant of the roll matrix R along its first row we get

$$\begin{aligned} \det(P) &= \cos^2 v + \sin^2 v = 1 \\ \det(R) &= \cos^2 w + \sin^2 w = 1 \end{aligned}$$

(b) If Y is a yaw matrix with yaw angle u then

$$Y^T = \begin{pmatrix} \cos u & -\sin u & 0 \\ \sin u & \cos u & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos(-u) & \sin(-u) & 0 \\ -\sin(-u) & \cos(-u) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

so Y^T is the matrix representing a yaw transformation with angle $-u$. It is easily verified that $Y^TY = I$ and hence that $Y^{-1} = Y^T$.

(c) By the same reasoning used in part (b) you can show that for the pitch matrix P and roll matrix R their inverses are their transposes. So if $Q = YPR$ then Q is nonsingular and

$$Q^{-1} = (YPR)^{-1} = R^{-1}P^{-1}Y^{-1} = R^TP^TY^T$$

14. (b) $\begin{pmatrix} 3/2 \\ -2 \end{pmatrix}$; (c) $\begin{pmatrix} 3/2 \\ 0 \end{pmatrix}$

16. If $L(\mathbf{x}) = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$ and A is the standard matrix representation of L, then $A\mathbf{x} = \mathbf{0}$. It follows from Theorem 1.4.2 that A is singular.

17. The proof is by induction on m. In the case that $m = 1$, $A^1 = A$ represents $L^1 = L$. If now A^k is the matrix representing L^k and if $\mathbf{x}$ is the coordinate vector of $\mathbf{v}$, then $A^k\mathbf{x}$ is the coordinate vector of $L^k(\mathbf{v})$. Since

$$L^{k+1}(\mathbf{v}) = L(L^k(\mathbf{v}))$$

it follows that

$$AA^k\mathbf{x} = A^{k+1}\mathbf{x}$$

is the coordinate vector of $L^{k+1}(\mathbf{v})$.

18. (b) $\begin{pmatrix} -5 & -2 & 4 \\ 3 & 2 & -2 \end{pmatrix}$

19. If $\mathbf{x} = [\mathbf{v}]_E$, then $A\mathbf{x} = [L_1(\mathbf{v})]_F$ and $B(A\mathbf{x}) = [L_2(L_1(\mathbf{v}))]_G$. Thus, for all $\mathbf{v} \in V$

$$(BA)[\mathbf{v}]_E = [L_2 \circ L_1(\mathbf{v})]_G$$

Hence BA is the matrix representing $L_2 \circ L_1$ with respect to E and G.

20. (a) Since A is the matrix representing L with respect to E and F, it follows that $L(\mathbf{v}) = \mathbf{0}_W$ if and only if $A[\mathbf{v}]_E = \mathbf{0}$. Thus $\mathbf{v} \in \ker(L)$ if and only if $[\mathbf{v}]_E \in N(A)$.

(b) Since A is the matrix representing L with respect to E and F, then it follows that $\mathbf{w} = L(\mathbf{v})$ if and only if $[\mathbf{w}]_F = A[\mathbf{v}]_E$. Thus, $\mathbf{w} \in L(V)$ if and only if $[\mathbf{w}]_F$ is in the column space of A.

SECTION 3

7. If A is similar to B then there exists a nonsingular matrix S_1 such that $A = S_1^{-1}BS_1$. Since B is similar to C there exists a nonsingular matrix S_2 such that $B = S_2^{-1}CS_2$. It follows that

$$A = S_1^{-1}BS_1 = S_1^{-1}S_2^{-1}CS_2S_1$$

If we set $S = S_2S_1$, then S is nonsingular and $S^{-1} = S_1^{-1}S_2^{-1}$. Thus $A = S^{-1}CS$ and hence A is similar to C.

8. (a) If $A = S\Lambda S^{-1}$, then $AS = \Lambda S$. If $\mathbf{s}_i$ is the ith column of S then $A\mathbf{s}_i$ is the ith column of AS and $\lambda_i\mathbf{s}_i$ is the ith column of ΛS. Thus

$$A\mathbf{s}_i = \lambda_i\mathbf{s}_i, \qquad i = 1, \ldots, n$$

(b) The proof is by induction on k. In the case $k = 1$ we have by part (a):

$$A\mathbf{x} = \alpha_1 A\mathbf{s}_1 + \cdots + \alpha_n A\mathbf{s}_n = \alpha_1\lambda_1\mathbf{s}_1 + \cdots + \alpha_n\lambda_n\mathbf{s}_n$$

If the result holds in the case $k = m$

$$A^m\mathbf{x} = \alpha_1\lambda_1^m\mathbf{s}_1 + \cdots + \alpha_n\lambda_n^m\mathbf{s}_n$$

then

$$\begin{aligned} A^{m+1}\mathbf{x} &= \alpha_1\lambda_1^m A\mathbf{s}_1 + \cdots + \alpha_n\lambda_n^m A\mathbf{s}_n \\ &= \alpha_1\lambda_1^{m+1}\mathbf{s}_1 + \cdots + \alpha_n\lambda_n^{m+1}\mathbf{s}_n \end{aligned}$$

Therefore by mathematical induction the result holds for all natural numbers k.

(c) If $|\lambda_i| < 1$ then $\lambda_i^k \to 0$ as $k \to \infty$. It follows from part (b) that $A^k\mathbf{x} \to \mathbf{0}$ as $k \to \infty$.

9. If $A = ST$ then

$$S^{-1}AS = S^{-1}STS = TS = B$$

Therefore B is similar to A.

10. If A and B are similar, then there exists a nonsingular matrix S such that

$$A = SBS^{-1}$$

If we set

$$T = BS^{-1}$$

then

$$A = ST \qquad \text{and} \qquad B = TS$$

11. If $B = S^{-1}AS$, then

$$\begin{aligned} \det(B) &= \det(S^{-1}AS) \\ &= \det(S^{-1})\det(A)\det(S) \\ &= \det(A) \end{aligned}$$

since

$$\det(S^{-1}) = \frac{1}{\det(S)}$$

12. (a) If $B = S^{-1}AS$, then

$$\begin{aligned} B^T &= (S^{-1}AS)^T \\ &= S^TA^T(S^{-1})^T \\ &= S^TA^T(S^T)^{-1} \end{aligned}$$

Therefore B^T is similar to A^T.

(b) If $B = S^{-1}AS$, then one can prove using mathematical induction that

$$B^k = S^{-1}A^kS$$

for any positive integer k. Therefore that B^k and A^k are similar for any positive integer k.

13. If A is similar to B and A is nonsingular, then

$$A = SBS^{-1}$$

and hence

$$B = S^{-1}AS$$

Since B is a product of nonsingular matrices it is nonsingular and

$$B^{-1} = (S^{-1}AS)^{-1} = S^{-1}A^{-1}S$$

Therefore B^{-1} and A^{-1} are similar.

14. If A and B are similar, then there exists a nonsingular matrix S such that $B = SAS^{-1}$.

(a) $A - \lambda I$ and $B - \lambda I$ are similar since

$$S(A - \lambda I)S^{-1} = SAS^{-1} - \lambda SIS^{-1} = B - \lambda I$$

(b) Since $A - \lambda I$ and $B - \lambda I$ are similar, it follows from Exercise 11 that their determinants are equal.

15. (a) Let $C = AB$ and $E = BA$. The diagonal entries of C and E are given by

$$c_{ii} = \sum_{k=1}^{n} a_{ik}b_{ki}, \qquad e_{kk} = \sum_{i=1}^{n} b_{ki}a_{ik}$$

Hence it follows that

$$\operatorname{tr}(AB) = \sum_{i=1}^{n} c_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik}b_{ki} = \sum_{k=1}^{n}\sum_{i=1}^{n} b_{ki}a_{ik} = \sum_{k=1}^{n} e_{kk} = \operatorname{tr}(BA)$$

(b) If B is similar to A, then $B = S^{-1}AS$. It follows from part (a) that

$$\operatorname{tr}(B) = \operatorname{tr}(S^{-1}(AS)) = \operatorname{tr}((AS)S^{-1}) = \operatorname{tr}(A)$$

MATLAB EXERCISES

2. (a) To determine the matrix representation of L with respect to E set

$$B = U^{-1}AU$$

(b) To determine the matrix representation of L with respect to F set

$$C = V^{-1}AV$$

(c) If B and C are both similar to A then they must be similar to each other. Indeed the transition matrix S from F to E is given by $S = U^{-1}V$ and

$$C = S^{-1}BS$$

CHAPTER TEST A

1. The statement is false in general. If $L : R^n \to R^m$ has matrix representation A and the rank of A is less than n, then it is possible to find vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ such that $L(\mathbf{x}_1) = L(\mathbf{x}_2)$ and $\mathbf{x}_1 \neq \mathbf{x}_2$. For example if

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}, \quad \mathbf{x}_1 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

and $L : R^2 \to R^2$ is defined by $L(\mathbf{x}) = A\mathbf{x}$, then

$$L(\mathbf{x}_1) = A\mathbf{x}_1 = \begin{pmatrix} 5 \\ 10 \end{pmatrix} = A\mathbf{x}_2 = L(\mathbf{x}_2)$$

2. The statement is true. If $\mathbf{v}$ is any vector in V and c is any scalar, then

$$\begin{aligned} (L_1 + L_2)(c\mathbf{v}) &= L_1(c\mathbf{v}) + L_2(c\mathbf{v}) \\ &= cL_1(\mathbf{v}) + cL_2(\mathbf{v}) \\ &= c(L_1(\mathbf{v}) + L_2(\mathbf{v})) \\ &= c(L_1 + L_2)(\mathbf{v}) \end{aligned}$$

If $\mathbf{v}_1$ and $\mathbf{v}_2$ are any vectors in V, then

$$\begin{aligned} (L_1 + L_2)(\mathbf{v}_1 + \mathbf{v}_2) &= L_1(\mathbf{v}_1 + \mathbf{v}_2) + L_2(\mathbf{v}_1 + \mathbf{v}_2) \\ &= L_1(\mathbf{v}_1) + L_1(\mathbf{v}_2) + L_2(\mathbf{v}_1) + L_2(\mathbf{v}_2) \\ &= (L_1(\mathbf{v}_1) + L_2(\mathbf{v}_1)) + (L_1(\mathbf{v}_2) + L_2(\mathbf{v}_2)) \\ &= (L_1 + L_2)(\mathbf{v}_1) + (L_1 + L_2)(\mathbf{v}_2) \end{aligned}$$

3. The statement is true. If $\mathbf{x}$ is in the kernel of L, then $L(\mathbf{x}) = \mathbf{0}$. Thus if $\mathbf{v}$ is any vector in V, then

$$L(\mathbf{v} + \mathbf{x}) = L(\mathbf{v}) + L(\mathbf{x}) = L(\mathbf{v}) + \mathbf{0} = L(\mathbf{v})$$

4. The statement is false in general. To see that $L_1 \neq L_2$, look at the effect of both operators on $\mathbf{e}_1$.

$$L_1(\mathbf{e}_1) = \begin{pmatrix} \frac{1}{2} \\ -\frac{\sqrt{3}}{2} \end{pmatrix} \qquad \text{and} \qquad L_2(\mathbf{e}_1) = \begin{pmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix}$$

5. The statement is false. The set of vectors in the homogeneous coordinate system does not form a subspace of R^3 since it is not closed under addition. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are vectors in the homogeneous system and $\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2$, then $\mathbf{y}$ is not a vector in the homogeneous coordinate system since $y_3 = 2$.

6. The statement is true. If A is the standard matrix representation of L, then

$$L^2(\mathbf{x}) = L(L(\mathbf{x})) = L(A\mathbf{x}) = A(A\mathbf{x}) = A^2\mathbf{x}$$

for any $\mathbf{x}$ in R^2. Clearly L^2 is a linear transformation since it can be represented by the matrix A^2.

7. The statement is true. If $\mathbf{x}$ is any vector in R^n then it can be represented in terms of the vectors of E

$$\mathbf{x} = c_1\mathbf{x}_{+}c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n$$

If L_1 and L_2 are both represented by the same matrix A with respect to E, then

$$L_1(\mathbf{x}) = d_1\mathbf{x}_+ d_2\mathbf{x}_2 + \cdots + d_n\mathbf{x}_n = L_2(\mathbf{x})$$

where $\mathbf{d} = A\mathbf{c}$. Since $L_1(\mathbf{x}) = L_2(\mathbf{x})$ for all $\mathbf{x} \in R^n$, it follows that $L_1 = L_2$.

8. The statement is true. See Theorem 4.3.1.

9. The statement is true. If A is similar to B and B is similar to C, then there exist nonsingular matrices X and Y such that

$$A = X^{-1}BX \qquad \text{and} \qquad B = Y^{-1}CY$$

If we set $Z = YX$, then Z is nonsingular and

$$A = X^{-1}BX = X^{-1}Y^{-1}CYX = Z^{-1}CZ$$

Thus A is similar to C.

10. The statement is false. Similar matrices have the same trace, but the converse is not true. For example, the matrices

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

have trace equal to 2, but the matrices not similar. In fact the only matrix that is similar to the identity matrix is I itself. (If S any nonsingular matrix, then $S^{-1}IS = I$.)

CHAPTER TEST B

1. (a) L is a linear operator since

$$L(c\mathbf{x}) = \begin{pmatrix} cx_1 + cx_2 \\ cx_1 \end{pmatrix} = c\begin{pmatrix} x_1 + x_2 \\ x_1 \end{pmatrix} = cL(\mathbf{x})$$

and

$$\begin{aligned} L(\mathbf{x}+\mathbf{y}) = \begin{pmatrix} (x_1+y_1)+(x_2+y_2) \\ x_1+y_1 \end{pmatrix} &= \begin{pmatrix} x_1+x_2 \\ x_1 \end{pmatrix} + \begin{pmatrix} y_1+y_2 \\ y_1 \end{pmatrix} \\ &= L(\mathbf{x}) + L(\mathbf{y}) \end{aligned}$$

(b) L is not a letter operator. If, for example we take $\mathbf{x} = (1,1)^T$ then

$$L(2\mathbf{x}) = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \qquad \text{and} \qquad 2L(\mathbf{x}) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

2. To determine the value of $L(\mathbf{v}_3)$ we must first express $\mathbf{v}_3$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. Thus we must find constants c_1 and c_2 such that $\mathbf{v}_3 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. In we set $V = (\mathbf{v}_1, \mathbf{v}_2)$ and solve the system $V\mathbf{c} = \mathbf{v}_3$ we see that $\mathbf{c} = (3,2)^T$. It follows then that

$$L(\mathbf{v}_3) = L(3\mathbf{v}_1 + 2\mathbf{v}_2) = 3L(\mathbf{v}_1) + 2L(\mathbf{v}_2) = \begin{pmatrix} 0 \\ 17 \end{pmatrix}$$

3. (a) $\ker(L) = \operatorname{Span}((1,1,1)^T)$

(b) $L(S) = \operatorname{Span}((-1,1,0)^T)$

4. Since

$$L(\mathbf{x}) = \begin{pmatrix} x_2 \\ x_1 \\ x_1 + x_2 \end{pmatrix} = x_1 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

it follows that the range of L is the span of the vectors

$$\mathbf{y}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{y}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

5. Let $\mathbf{e}_1$ and $\mathbf{e}_2$ be the standard basis vectors for R^2. To determine the matrix representation of L we set

$$\mathbf{a}_1 = L(\mathbf{e}_1) = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}, \quad \mathbf{a}_2 = L(\mathbf{e}_2) = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$$

If we set

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 3 & 2 \end{pmatrix}$$

then $L(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in R^2$.

6. To determine the matrix representation we set

$$\mathbf{a}_1 = L(\mathbf{e}_1) = \begin{pmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix} \qquad \text{and} \qquad \mathbf{a}_2 = L(\mathbf{e}_2) = \begin{pmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix}$$

The matrix representation of the operator is

$$A = (\mathbf{a}_1, \mathbf{a}_2) = \begin{pmatrix} -\frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$$

7. $A = \begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$

8. The standard matrix representation for a 45° counterclockwise rotation operator is

$$A = \begin{pmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

The matrix representation with respect to the basis $[\mathbf{u}_1, \mathbf{u}_2]$ is

$$B = U^{-1}AU = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} -\frac{16}{\sqrt{2}} & -\frac{29}{\sqrt{2}} \\ \frac{10}{\sqrt{2}} & \frac{18}{\sqrt{2}} \end{pmatrix}$$

9. **(a)** If $U = (\mathbf{u}_1, \mathbf{u}_2)$ and $V = (\mathbf{v}_1, \mathbf{v}_2)$ then the transition matrix S from $[\mathbf{v}_1, \mathbf{v}_2]$ to $[\mathbf{u}_1, \mathbf{u}_2]$ is

$$S = U^{-1}V = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} 12 & 7 \\ -7 & -4 \end{pmatrix}$$

(b) By Theorem 4.3.1 the matrix representation of L with respect to $[\mathbf{v}_1, \mathbf{v}_2]$ is

$$B = S^{-1}AS = \begin{pmatrix} -4 & -7 \\ 7 & 12 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 12 & 7 \\ -7 & -4 \end{pmatrix} = \begin{pmatrix} -222 & -131 \\ 383 & 226 \end{pmatrix}$$

10. (a) If A and B are similar then $B = S^{-1}AS$ for some nonsingular matrix S. It follows then that

$$\begin{aligned} \det(B) = \det(S^{-1}AS) &= \det(S^{-1})\det(A)\det(S) \\ &= \frac{1}{\det(S)}\det(A)\det(S) = \det(A) \end{aligned}$$

(b) If $B = S^{-1}AS$ then

$$S^{-1}(A - \lambda I)S = S^{-1}AS - \lambda S^{-1}IS = B - \lambda I$$

Therefore $A - \lambda I$ and $B - \lambda I$ ate similar and it follows from part (a) that their determinants must be equal.

Chapter 5

SECTION 1

1. (c) $\cos\theta = \dfrac{14}{\sqrt{221}}, \qquad \theta \approx 10.65°$

(d) $\cos\theta = \dfrac{4\sqrt{6}}{21}, \qquad \theta \approx 62.19°$

3. (b) $\mathbf{p} = (4,4)^T,\ \ \mathbf{x}-\mathbf{p} = (-1,1)^T$

$\mathbf{p}^T(\mathbf{x}-\mathbf{p}) = -4+4 = 0$

(d) $\mathbf{p} = (-2,-4,2)^T,\ \ \mathbf{x}-\mathbf{p} = (4,-1,2)^T$

$\mathbf{p}^T(\mathbf{x}-\mathbf{p}) = -8+4+4 = 0$

4. If $\mathbf{x}$ and $\mathbf{y}$ are linearly independent and θ is the angle between the vectors, then $|\cos\theta| < 1$ and hence

$$|\mathbf{x}^T\mathbf{y}| = \|\mathbf{x}\|\,\|\mathbf{y}\|\,|\cos\theta| < 6$$

8. (b) $-3(x-4)+6(y-2)+2(z+5)=0$

11. (a) $\mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2 \geq 0$

(b) $\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 = y_1x_1 + y_2x_2 = \mathbf{y}^T\mathbf{x}$

(c) $\mathbf{x}^T(\mathbf{y}+\mathbf{z}) = x_1(y_1+z_1) + x_2(y_2+z_2)$
$= (x_1y_1 + x_2y_2) + (x_1z_2 + x_2z_2)$
$= \mathbf{x}^T\mathbf{y} + \mathbf{x}^T\mathbf{z}$

12. The inequality can be proved using the Cauchy-Schwarz inequality as follows:

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= (\mathbf{u}+\mathbf{v})^T(\mathbf{u}+\mathbf{v})\\ &= \mathbf{u}^T\mathbf{u} + \mathbf{v}^T\mathbf{u} + \mathbf{u}^T\mathbf{v} + \mathbf{v}^T\mathbf{v}\\ &= \|\mathbf{u}\|^2 + 2\mathbf{u}^T\mathbf{v} + \|\mathbf{v}\|^2\\ &= \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\,\|\mathbf{v}\|\,\cos\theta + \|\mathbf{v}\|^2\\ &\le \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2\\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2\end{aligned}$$

Taking square roots, we get

$$\|\,\mathbf{u}+\mathbf{v}\,\| \le \|\,\mathbf{u}\,\| + \|\,\mathbf{v}\,\|$$

Equality will hold if and only if $\cos\theta = 1$. This will happen if one of the vectors is a multiple of the other. Geometrically one can think of $\|\mathbf{u}\|$ and $\|\mathbf{v}\|$ as representing the lengths of two sides of a triangle. The length of the third side of the triangle will be $\|\mathbf{u}+\mathbf{v}\|$. Clearly the length of the third side must be less than the sum of the lengths of the first two sides. In the case of equality the triangle degenerates to a line segment.

13. No. For example, if $\mathbf{x}_1 = \mathbf{e}_1$, $\mathbf{x}_2 = \mathbf{e}_2$, $\mathbf{x}_3 = 2\mathbf{e}_1$, then $\mathbf{x}_1 \perp \mathbf{x}_2$, $\mathbf{x}_2 \perp \mathbf{x}_3$, but $\mathbf{x}_1$ is not orthogonal to $\mathbf{x}_3$.

14. (a) By the Pythagorean Theorem

$$\alpha^2 + h^2 = \|\mathbf{a}_1\|^2$$

where α is the scalar projection of $\mathbf{a}_1$ onto $\mathbf{a}_2$. It follows that

$$\alpha^2 = \frac{(\mathbf{a}_1^T\mathbf{a}_2)^2}{\|\mathbf{a}_2\|^2}$$

and

$$h^2 = \|\mathbf{a}_1\|^2 - \frac{(\mathbf{a}_1^T\mathbf{a}_2)^2}{\|\mathbf{a}_2\|^2}$$

Hence

$$h^2\|\mathbf{a}_2\|^2 = \|\mathbf{a}_1\|^2\,\|\mathbf{a}_2\|^2 - (\mathbf{a}_1^T\mathbf{a}_2)^2$$

(b) If $\mathbf{a}_1 = (a_{11}, a_{21})^T$ and $\mathbf{a}_2 = (a_{12}, a_{22})^T$, then by part (a)

$$\begin{aligned}h^2\|\mathbf{a}_2\|^2 &= (a_{11}^2 + a_{21}^2)(a_{12}^2 + a_{22}^2) - (a_{11}a_{12} + a_{21}a_{22})^2\\ &= (a_{11}^2a_{22}^2 - 2a_{11}a_{22}a_{12}a_{21} + a_{21}^2a_{12}^2)\\ &= (a_{11}a_{22} - a_{21}a_{12})^2\end{aligned}$$

Therefore

$$\text{Area of } P = h\|\mathbf{a}_2\| = |a_{11}a_{22} - a_{21}a_{12}| = |\det(A)|$$

15. (a) It θ is the angle between $\mathbf{x}$ and $\mathbf{y}$, then

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|} = \frac{20}{8\cdot 5} = \frac{1}{2}, \quad \theta = \frac{\pi}{3}$$

(b) The distance between the vectors is given by

$$\|\mathbf{x}-\mathbf{y}\| = \sqrt{0^2+2^2+(-6)^2+3^2} = 7$$

16. (a) Let

$$\alpha = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \qquad \text{and} \qquad \beta = \frac{(\mathbf{x}^T\mathbf{y})^2}{\mathbf{y}^T\mathbf{y}}$$

In terms of these scalars we have $\mathbf{p} = \alpha\mathbf{y}$ and $\mathbf{p}^T\mathbf{x} = \beta$. Furthermore

$$\mathbf{p}^T\mathbf{p} = \alpha^2\mathbf{y}^T\mathbf{y} = \beta$$

and hence

$$\mathbf{p}^T\mathbf{z} = \mathbf{p}^T\mathbf{x} - \mathbf{p}^T\mathbf{p} = \beta - \beta = 0$$

(b) If $\|\mathbf{p}\| = 6$ and $\|\mathbf{z}\| = 8$, then we can apply the Pythagorean law to determine the length of $\mathbf{x} = \mathbf{p} + \mathbf{z}$. It follows that

$$\|\mathbf{x}\|^2 = \|\mathbf{p}\|^2 + \|\mathbf{z}\|^2 = 36 + 64 = 100$$

and hence $\|\mathbf{x}\| = 10$.

17. The matrix Q is unchanged and the nonzero entries of our new search vector $\mathbf{x}$ are $x_6 = \frac{\sqrt{6}}{3}$, $x_7 = \frac{\sqrt{6}}{6}$, $x_{10} = \frac{\sqrt{6}}{6}$. Rounded to three decimal places the search vector is

$$\mathbf{x} = (0,\ 0,\ 0,\ 0,\ 0,\ 0.816,\ 0.408,\ 0,\ 0,\ 0.408)^T$$

The search results are given by the vector

$$\mathbf{y} = Q^T\mathbf{x} = (0,\ 0.161,\ 0.401\, 0.234,\ 0.612,\ 0.694,\ 0,\ 0.504)^T$$

The largest entry of $\mathbf{y}$ is $y_6 = 0.694$. This implies that Module 6 is the one that best meets our search criteria.

SECTION 2

1. (b) The reduced row echelon form of A is

$$\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \end{pmatrix}$$

The set $\{(2,\ -1,\ 1)^T\}$ is a basis for $N(A)$ and $\{(1,\ 0,\ -2)^T,\ (0,\ 1,\ 1)^T\}$ is a basis for $R(A^T)$. The reduced row echelon form of A^T is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$N(A^T) = \{(0,\ 0)^T\}$ and $\{(1,\ 0)^T,\ (0,\ 1)^T\}$ is a basis for $R(A) = R^2$.

(c) The reduced row echelon form of A is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$N(A) = \{(0,\ 0)^T\}$ and $\{(1,\ 0)^T,\ (0,\ 1)^T\}$ is a basis for $R(A^T)$. The reduced row echelon form of A^T is

$$U = \begin{pmatrix} 1 & 0 & \frac{5}{14} & \frac{5}{14} \\ 0 & 1 & \frac{4}{7} & \frac{11}{7} \end{pmatrix}$$

We can obtain a basis for $R(A)$ by transposing the rows of U and we can obtain a basis for $N(A^T)$ by solving $U\mathbf{x} = \mathbf{0}$. It follows that

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ \frac{5}{14} \\ \frac{5}{14} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \frac{4}{7} \\ \frac{11}{7} \end{pmatrix} \right\} \quad \text{and} \quad \left\{ \begin{pmatrix} -\frac{5}{14} \\ -\frac{4}{7} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{5}{14} \\ -\frac{11}{7} \\ 0 \\ 1 \end{pmatrix} \right\}$$

are bases for $R(A)$ and $N(A^T)$, respectively.

2. (b) S corresponds to a line ℓ in 3-space that passes through the origin and the point $(1,\ -1,\ 1)$. $S^\perp$ corresponds to a plane in 3-space that passes through the origin and is normal to the line ℓ.

3. (a) A vector $\mathbf{z}$ will be in $S^\perp$ if and only if $\mathbf{z}$ is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$. Since $\mathbf{x}^T$ and $\mathbf{y}^T$ are the row vectors of A, it follows that $S^\perp = N(A)$.

6. No. $(3,\ 1,\ 2)^T$ and $(2,\ 1,\ 1)^T$ are not orthogonal.

7. No. Since $N(A^T)$ and $R(A)$ are orthogonal complements

$$N(A^T) \cap R(A) = \{\mathbf{0}\}$$

The vector $\mathbf{a}_j$ cannot be in $N(A^T)$ since it is a nonzero element of $R(A)$. Also, note that the jth coordinate of $A^T\mathbf{a}_j$ is

$$\mathbf{a}_j^T\mathbf{a}_j = \|\mathbf{a}_j\|^2 > 0$$

8. If $\mathbf{y} \in S^\perp$ then since each $\mathbf{x}_i \in S$ it follows that $\mathbf{y} \perp \mathbf{x}_i$ for $i = 1, \ldots, k$. Conversely if $\mathbf{y} \perp \mathbf{x}_i$ for $i = 1, \ldots, k$ and $\mathbf{x} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_k\mathbf{x}_k$ is any element of S, then

$$\mathbf{y}^T\mathbf{x} = \mathbf{y}^T \left(\sum_{i=1}^{k} \alpha_i\mathbf{x}_i \right) = \sum_{i=1}^{k} \alpha_i\mathbf{y}^T\mathbf{x}_i = 0$$

Thus $\mathbf{y} \in S^\perp$.

10. **Corollary 5.2.5.** If A is an $m \times n$ matrix and $\mathbf{b} \in R^n$, then either there is a vector $\mathbf{x} \in R^n$ such that $A\mathbf{x} = \mathbf{b}$ or there is a vector $\mathbf{y} \in R^m$ such that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$.

Proof: If $A\mathbf{x} = \mathbf{b}$ has no solution then $\mathbf{b} \notin R(A)$. Since $R(A) = N(A^T)^\perp$ it follows that $\mathbf{b} \notin N(A^T)^\perp$. But this means that there is a vector $\mathbf{y}$ in $N(A^T)$ that is not orthogonal to $\mathbf{b}$. Thus $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$.

11. If $\mathbf{x}$ is not a solution to $A\mathbf{x} = \mathbf{0}$ then $\mathbf{x} \notin N(A)$. Since $N(A) = R(A^T)^\perp$ it follows that $\mathbf{x} \notin R(A^T)^\perp$. Thus there exists a vector $\mathbf{y}$ in $R(A^T)$ that is not orthogonal to $\mathbf{x}$, i.e., $\mathbf{x}^T\mathbf{y} \neq \mathbf{0}$.

12. Part (a) follows since $R^n = N(A) \oplus R(A^T)$.
Part (b) follows since $R^m = N(A^T) \oplus R(A)$.

13. (a) $A\mathbf{x} \in R(A)$ for all vectors $\mathbf{x}$ in R^n. If $\mathbf{x} \in N(A^TA)$ then

$$A^TA\mathbf{x} = \mathbf{0}$$

and hence $A\mathbf{x} \in N(A^T)$.

(b) If $\mathbf{x} \in N(A)$, then

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

and hence $\mathbf{x} \in N(A^TA)$. Thus $N(A)$ is a subspace of $N(A^TA)$.

Conversely, if $\mathbf{x} \in N(A^TA)$, then by part (a), $A\mathbf{x} \in R(A) \cap N(A^T)$. Since $R(A) \cap N(A^T) = \{\mathbf{0}\}$, it follows that $\mathbf{x} \in N(A)$. Thus $N(A^TA)$ is a subspace of $N(A)$. It follows then that $N(A^TA) = N(A)$.

(c) A and A^TA have the same nullspace and consequently must have the same nullity. Since both matrices have n columns, it follows from the Rank-Nullity Theorem that they must also have the same rank.

(d) If A has linearly independent columns then A has rank n. By part (c), A^TA also has rank n and consequently is nonsingular.

14. (a) If $\mathbf{x} \in N(B)$, then

$$C\mathbf{x} = AB\mathbf{x} = A\mathbf{0} = \mathbf{0}$$

Thus $\mathbf{x} \in N(C)$ and it follows that $N(B)$ is a subspace of $N(C)$.

(b) If $\mathbf{x} \in N(C)^\perp$, then $\mathbf{x}^T\mathbf{y} = 0$ for all $\mathbf{y} \in N(C)$. Since $N(B) \subset N(C)$ it follows that $\mathbf{x}$ is orthogonal to each element of $N(B)$ and hence $\mathbf{x} \in N(B)^\perp$. Therefore

$$R(C^T) = N(C)^\perp \text{ is a subspace of } N(B)^\perp = R(B^T)$$

15. Let $\mathbf{x} \in U \cap V$. We can write

$$\begin{aligned} \mathbf{x} &= \mathbf{0} + \mathbf{x} \qquad (\mathbf{0} \in U, \quad \mathbf{x} \in V) \\ \mathbf{x} &= \mathbf{x} + \mathbf{0} \qquad (\mathbf{x} \in U, \quad \mathbf{0} \in V) \end{aligned}$$

By the uniqueness of the direct sum representation $\mathbf{x} = \mathbf{0}$.

16. It was shown in the text that

$$R(A) = \{A\mathbf{y} \mid \mathbf{y} \in R(A^T)\}$$

If $\mathbf{y} \in R(A^T)$, then we can write

$$\mathbf{y} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_r\mathbf{x}_r$$

Thus

$$A\mathbf{y} = \alpha_1 A\mathbf{x}_1 + \alpha_2 A\mathbf{x}_2 + \cdots + \alpha_r A\mathbf{x}_r$$

and it follows that the vectors $A\mathbf{x}_1, \ldots, A\mathbf{x}_r$ span $R(A)$. Since $R(A)$ has dimension r, $\{A\mathbf{x}_1, \ldots, A\mathbf{x}_r\}$ is a basis for $R(A)$.

17. **(a)** A is symmetric since

$$\begin{aligned} A^T &= (\mathbf{x}\mathbf{y}^T + \mathbf{y}\mathbf{x}^T)^T = (\mathbf{x}\mathbf{y}^T)^T + (\mathbf{y}\mathbf{x}^T)^T \\ &= (\mathbf{y}^T)^T\mathbf{x}^T + (\mathbf{x}^T)^T\mathbf{y}^T = \mathbf{y}\mathbf{x}^T + \mathbf{x}\mathbf{y}^T = A \end{aligned}$$

(b) For any vector $\mathbf{z}$ in R^n

$$A\mathbf{z} = \mathbf{x}\mathbf{y}^T\mathbf{z} + \mathbf{y}\mathbf{x}^T\mathbf{z} = c_1\mathbf{x} + c_2\mathbf{y}$$

where $c_1 = \mathbf{y}^T\mathbf{z}$ and $c_2 = \mathbf{x}^T\mathbf{z}$. If $\mathbf{z}$ is in $N(A)$ then

$$\mathbf{0} = A\mathbf{z} = c_1\mathbf{x} + c_2\mathbf{y}$$

and since $\mathbf{x}$ and $\mathbf{y}$ are linearly independent we have $\mathbf{y}^T\mathbf{z} = c_1 = 0$ and $\mathbf{x}^T\mathbf{z} = c_2 = 0$. So $\mathbf{z}$ is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$. Since $\mathbf{x}$ and $\mathbf{y}$ span S it follows that $\mathbf{z} \in S^\perp$.

Conversely, if $\mathbf{z}$ is in $S^\perp$ then $\mathbf{z}$ is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$. It follows that

$$A\mathbf{z} = c_1\mathbf{x} + c_2\mathbf{y} = \mathbf{0}$$

since $c_1 = \mathbf{y}^T\mathbf{z} = 0$ and $c_2 = \mathbf{x}^T\mathbf{z} = 0$. Therefore $\mathbf{z}$ is in $N(A)$ and hence $N(A) = S^\perp$.

(c) Clearly $\dim S = 2$ and by Theorem 5.2.2, $\dim S + \dim S^\perp = n$. Using our result from part (a) we have

$$\dim N(A) = \dim S^\perp = n - 2$$

So A has nullity $n - 2$. It follows from the Rank-Nullity Theorem that the rank of A must be 2.

SECTION 3

1. (b) $A^TA = \begin{pmatrix} 6 & -1 \\ -1 & 6 \end{pmatrix}$ and $A^T\mathbf{b} = \begin{pmatrix} 20 \\ -25 \end{pmatrix}$

The solution to the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$ is

$$\mathbf{x} = \begin{pmatrix} 19/7 \\ -26/7 \end{pmatrix}$$

2. (Exercise 1b.)

(a) $\mathbf{p} = \frac{1}{7}(-45,\ 12,\ 71)^T$

(b) $\mathbf{r} = \frac{1}{7}(115,\ 23,\ 69)^T$

(c)

$$A^T\mathbf{r} = \begin{pmatrix} -1 & 2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \begin{pmatrix} \frac{115}{7} \\ \frac{23}{7} \\ \frac{69}{7} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Therefore $\mathbf{r}$ is in $N(A^T)$.

6. $A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 9 \end{pmatrix}$

$A^T A = \begin{pmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{pmatrix}, \quad A^T\mathbf{b} = \begin{pmatrix} 13 \\ 21 \\ 39 \end{pmatrix}$

The solution to $A^T A\mathbf{x} = A^T\mathbf{b}$ is $(0.6,\ 1.7,\ 1.2)^T$. Therefore the best least squares fit by a quadratic polynomial is given by

$$p(x) = 0.6 + 1.7x + 1.2x^2$$

7. To find the best fit by a linear function we must find the least squares solution to the linear system

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

If we form the normal equations the augmented matrix for the system will be

$$\left(\begin{array}{cc|c} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i \end{array}\right)$$

If $\overline{x} = 0$ then

$$\sum_{i=1}^{n} x_i = n\overline{x} = 0$$

and hence the coefficient matrix for the system is diagonal. The solution is easily obtained.

$$c_0 = \frac{\sum_{i=1}^{n} y_i}{n} = \overline{y}$$

and

$$c_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x}}$$

8. To show that the least squares line passes through the center of mass, we introduce a new variable $z = x - \overline{x}$. If we set $z_i = x_i - \overline{x}$ for $i = 1, \ldots, n$, then $\overline{z} = 0$. Using the result from Exercise 7 the equation of the best least squares fit by a linear function in the new zy-coordinate system is

$$y = \overline{y} + \frac{\mathbf{z}^T\mathbf{y}}{\mathbf{z}^T\mathbf{z}} z$$

If we translate this back to xy-coordinates we end up with the equation

$$y - \overline{y} = c_1(x - \overline{x})$$

where

$$c_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

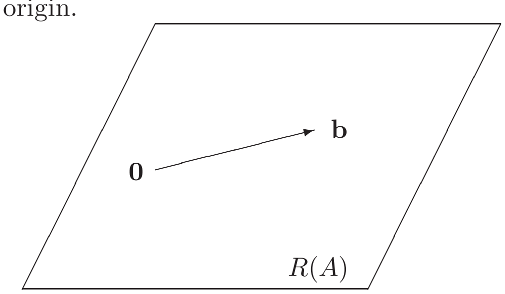
9. (a) If $\mathbf{b} \in R(A)$ then $\mathbf{b} = A\mathbf{x}$ for some $\mathbf{x} \in R^n$. It follows that

$$P\mathbf{b} = PA\mathbf{x} = A(A^TA)^{-1}A^TA\mathbf{x} = A\mathbf{x} = \mathbf{b}$$

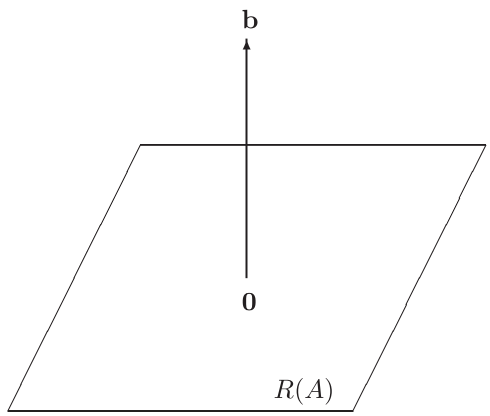
(b) If $\mathbf{b} \in R(A)^{\perp}$ then since $R(A)^{\perp} = N(A^T)$ it follows that $A^T\mathbf{b} = \mathbf{0}$ and hence

$$P\mathbf{b} = A(A^TA)^{-1}A^T\mathbf{b} = \mathbf{0}$$

(c) The following figures give a geometric illustration of parts (a) and (b). In the first figure $\mathbf{b}$ lies in the plane corresponding to $R(A)$. Since it is already in the plane, projecting it onto the plane will have no effect. In the second figure $\mathbf{b}$ lies on the line through the origin that is normal to the plane. When it is projected onto the plane it projects right down to the origin.



If $\mathbf{b} \in R(A)$, then $P\mathbf{b} = \mathbf{b}$.



If $\mathbf{b} \in R(A)^{\perp}$, then $P\mathbf{b} = \mathbf{0}$.

10. (a) By the Consistency Theorem $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is in $R(A)$. We are given that $\mathbf{b}$ is in $N(A^T)$. So if the system is consistent then $\mathbf{b}$ would be in $R(A) \cap N(A^T) = \{\mathbf{0}\}$. Since $\mathbf{b} \neq \mathbf{0}$, the system must be inconsistent.

(b) If A has rank 3 then A^TA also has rank 3 (see Exercise 13 in Section 2). The normal equations are always consistent and in this case there will be 2 free variables. So the least squares problem will have infinitely many solutions.

11. (a) $P^2 = A(A^TA)^{-1}A^TA(A^TA)^{-1}A^T = A(A^TA)^{-1}A^T = P$

(b) Prove: $P^k = P$ for $k = 1, 2, \ldots$.
Proof: The proof is by mathematical induction. In the case $k = 1$ we have $P^1 = P$. If $P^m = P$ for some m then

$$P^{m+1} = PP^m = PP = P^2 = P$$

(c)
$$\begin{aligned} P^T &= [A(A^TA)^{-1}A^T]^T \\ &= (A^T)^T[(A^TA)^{-1}]^TA^T \\ &= A[(A^TA)^T]^{-1}A^T \\ &= A(A^TA)^{-1}A^T \\ &= P \end{aligned}$$

12. If

$$\begin{pmatrix} A & I \\ O & A^T \end{pmatrix} \begin{pmatrix} \hat{\mathbf{x}} \\ \mathbf{r} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix}$$

then

$$\begin{aligned} A\hat{\mathbf{x}} + \mathbf{r} &= \mathbf{b} \\ A^T\mathbf{r} &= \mathbf{0} \end{aligned}$$

We have then that

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

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

Therefore

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

So $\hat{\mathbf{x}}$ is a solution to the normal equations and hence is the least squares solution to $A\mathbf{x} = \mathbf{b}$.

13. If $\hat{\mathbf{x}}$ is a solution to the least squares problem, then $\hat{\mathbf{x}}$ is a solution to the normal equations

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

It follows that a vector $\mathbf{y} \in R^n$ will be a solution if and only if

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

for some $\mathbf{z} \in N(A^TA)$. (See Exercise 20, Chapter 3, Section 6). Since

$$N(A^TA) = N(A)$$

we conclude that $\mathbf{y}$ is a least squares solution if and only if

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

for some $\mathbf{z} \in N(A)$.

SECTION 4

2. (b) $\mathbf{p} = \dfrac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y} = \dfrac{12}{72}\mathbf{y} = \left(\dfrac{4}{3}, \dfrac{1}{3}, \dfrac{1}{3}, 0\right)^T$

(c) $\mathbf{x} - \mathbf{p} = \left(-\dfrac{1}{3}, \dfrac{2}{3}, \dfrac{2}{3}, 1\right)^T$

$$(\mathbf{x} - \mathbf{p})^T\mathbf{p} = -\frac{4}{9} + \frac{2}{9} + \frac{2}{9} + 0 = 0$$

(d) $\|\mathbf{x} - \mathbf{p}\|_2 = \sqrt{2}$, $\|\mathbf{p}\|_2 = \sqrt{2}$, $\|\mathbf{x}\|_2 = 2$

$$\|\mathbf{x} - \mathbf{p}\|^2 + \|\mathbf{p}\|^2 = 4 = \|\mathbf{x}\|^2$$

3. (a) $\langle \mathbf{x}, \mathbf{y}\rangle = x_1y_1w_1 + x_2y_2w_2 + x_3y_3w_3 = 1 \cdot -5 \cdot \dfrac{1}{4} + 1 \cdot 1 \cdot \dfrac{1}{2} + 1 \cdot 3 \cdot \dfrac{1}{4} = 0$

5. (i)

$$\langle A, A\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 \ge 0$$

and $\langle A, A\rangle = 0$ if and only if each $a_{ij} = 0$.

(ii) $\langle A, B\rangle = \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij}a_{ij} = \langle B, A\rangle$

(iii)

$$\begin{aligned}\langle \alpha A + \beta B, C\rangle &= \sum_{i=1}^{m}\sum_{j=1}^{n}(\alpha a_{ij} + \beta b_{ij})c_{ij} \\ &= \alpha \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}c_{ij} + \beta \sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij}c_{ij} \\ &= \alpha\langle A, C\rangle + \beta\langle B, C\rangle\end{aligned}$$

6. Show that the inner product on $C[a,b]$ determined by

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx$$

satisfies the last two conditions of the definition of an inner product.

Solution:

(ii) $\langle f, g\rangle = \int_a^b f(x)g(x)\,dx = \int_a^b g(x)f(x)\,dx = \langle g, f\rangle$

(iii) $$\begin{aligned}\langle \alpha f + \beta g, h\rangle &= \int_a^b (\alpha f(x) + \beta g(x))h(x)\,dx \\ &= \alpha \int_a^b f(x)h(x)\,dx + \beta \int_a^b g(x)h(x)\,dx \\ &= \alpha\langle f, h\rangle + \beta\langle g, h\rangle\end{aligned}$$

7 (c)

$$\left\langle x^2, x^3\right\rangle = \int_0^1 x^2x^3dx = \frac{1}{6}$$

8 (c)

$$\|1\|^2 = \int_0^1 1\cdot 1\ dx = 1$$

$$\|\mathbf{p}\|^2 = \int_0^1 \frac{9}{4}x^2\,dx = \frac{3}{4}$$

$$\|1 - \mathbf{p}\|^2 = \int_0^1 \left(1 - \frac{3}{2}x\right)^2 dx = \frac{1}{4}$$

Thus $\|1\| = 1$, $\|\mathbf{p}\| = \frac{\sqrt{3}}{2}$, $\|1 - \mathbf{p}\| = \frac{1}{2}$, and

$$\|1 - \mathbf{p}\|^2 + \|\mathbf{p}\|^2 = 1 = \|1\|^2$$

9. The vectors $\cos mx$ and $\sin nx$ are orthogonal since

$$\begin{aligned}\langle \cos mx, \sin nx\rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos mx \sin nx\,dx \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} [\sin(n+m)x + \sin(n-m)x]\,dx \\ &= 0\end{aligned}$$

They are unit vectors since

$$\begin{aligned}\langle \cos mx, \cos mx\rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos^2 mx\,dx \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}[1+\cos 2mx]\,dx \\ &= 1\end{aligned}$$

$$\begin{aligned}\langle \sin nx, \sin nx\rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \sin nx \sin nx\,dx \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}(1-\cos 2nx)\,dx \\ &= 1\end{aligned}$$

Since the $\cos mx$ and $\sin nx$ are orthogonal, the distance between the vectors can be determined using the Pythagorean law.

$$\|\cos mx - \sin nx\| = (\|\cos mx\|^2 + \|\sin nx\|^2)^{\frac{1}{2}} = \sqrt{2}$$

10. $\langle x, x^2\rangle = \sum_{i=1}^{5} x_i x_i^2 = -1 - \frac{1}{8} + 0 + \frac{1}{8} + 1 = 0$

11. (c) $\|x - x^2\| = \left(\sum_{i=1}^{5}(x_i - x_i^2)^2\right)^{1/2} = \frac{\sqrt{26}}{4}$

12. (i) By the definition of an inner product we have $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$ with equality if and only if $\mathbf{v} = \mathbf{0}$. Thus $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle} \geq 0$ and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

(ii) $\|\alpha\mathbf{v}\| = \sqrt{\langle \alpha\mathbf{v}, \alpha\mathbf{v}\rangle} = \sqrt{\alpha^2\langle \mathbf{v}, \mathbf{v}\rangle} = |\alpha|\,\|\mathbf{v}\|$

13. (i) Clearly

$$\sum_{i=1}^{n} |x_i| \geq 0$$

If

$$\sum_{i=1}^{n} |x_i| = 0$$

then all of the x_i's must be 0.

(ii) $\|\alpha\mathbf{x}\|_1 = \sum_{i=1}^{n}|\alpha x_i| = |\alpha|\sum_{i=1}^{n}|x_i| = |\alpha|\,\|\mathbf{x}\|_1$

(iii) $\|\mathbf{x}+\mathbf{y}\|_1 = \sum_{i=1}^{n}|x_i + y_i| \leq \sum_{i=1}^{n}|x_i| + \sum_{i=1}^{n}|y_i| = \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1$

14. (i) $\|\mathbf{x}\|_\infty = \max_{1\leq i\leq n}|x_i| \geq 0$. If $\max_{1\leq i\leq n}|x_i| = 0$ then all of the x_i's must be zero.

(ii) $\|\alpha\mathbf{x}\|_\infty = \max_{1\leq i\leq n}|\alpha x_i| = |\alpha|\max_{1\leq i\leq n}|x_i| = |\alpha|\,\|\mathbf{x}\|_\infty$

(iii) $\|\mathbf{x}+\mathbf{y}\|_\infty = \max|x_i + y_i| \leq \max|x_i| + \max|y_i| = \|\mathbf{x}\|_\infty + \|\mathbf{y}\|_\infty$

17. If $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, then

$$\begin{aligned} \|\mathbf{x}-\mathbf{y}\|^2 &= \langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y} \rangle \\ &= \langle \mathbf{x}, \mathbf{x} \rangle - 2\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \end{aligned}$$

Therefore

$$\|\mathbf{x}-\mathbf{y}\| = (\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2)^{1/2}$$

Alternatively, one can prove this result by noting that if $\mathbf{x}$ is orthogonal to $\mathbf{y}$ then $\mathbf{x}$ is also orthogonal to $-\mathbf{y}$ and hence by the Pythagorean Law

$$\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}+(-\mathbf{y})\|^2 = \|\mathbf{x}\|^2 + \|-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

18. $\|\mathbf{x}-\mathbf{y}\| = (\langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y} \rangle)^{1/2} = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2}$

19. For $i = 1, \ldots, n$

$$|x_i| \le (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2} = \|\mathbf{x}\|_2$$

Thus

$$\|\mathbf{x}\|_\infty = \max_{1\le i\le n} |x_i| \le \|\mathbf{x}\|_2$$

20. $$\begin{aligned} \|\mathbf{x}\|_2 &= \|x_1\mathbf{e}_1 + x_2\mathbf{e}_2\|_2 \\ &\le \|x_1\mathbf{e}_1\|_2 + \|x_2\mathbf{e}_2\|_2 \\ &= |x_1|\,\|\mathbf{e}_1\|_2 + |x_2|\,\|\mathbf{e}_2\|_2 \\ &= |x_1| + |x_2| \\ &= \|\mathbf{x}\|_1 \end{aligned}$$

21. $\mathbf{e}_1$ and $\mathbf{e}_2$ are both examples.

22. $\|-\mathbf{v}\| = \|(-1)\mathbf{v}\| = |-1|\,\|\mathbf{v}\| = \|\mathbf{v}\|$

23. $$\begin{aligned} \|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle \\ &= \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 \\ &\ge \|\mathbf{u}\|^2 - 2\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\ &= (\|\mathbf{u}\| - \|\mathbf{v}\|)^2 \end{aligned}$$

24.

$$\begin{aligned} \|\mathbf{u}+\mathbf{v}\|^2 &= \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 \\ \|\mathbf{u}-\mathbf{v}\|^2 &= \|\mathbf{u}\|^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 \\ \|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 &= 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2 = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2) \end{aligned}$$

If the vectors $\mathbf{u}$ and $\mathbf{v}$ are used to form a parallelogram in the plane, then the diagonals will be $\mathbf{u}+\mathbf{v}$ and $\mathbf{u}-\mathbf{v}$. The equation shows that the sum of the squares of the lengths of the diagonals is twice the sum of the squares of the lengths of the two sides.

25. The result will not be valid for most choices of $\mathbf{u}$ and $\mathbf{v}$. For example, if $\mathbf{u} = \mathbf{e}_1$ and $\mathbf{v} = \mathbf{e}_2$, then

$$\begin{aligned} \|\mathbf{u}+\mathbf{v}\|_1^2 + \|\mathbf{u}-\mathbf{v}\|_1^2 &= 2^2 + 2^2 = 8 \\ 2\|\mathbf{u}\|_1^2 + 2\|\mathbf{v}\|_1^2 &= 2 + 2 = 4 \end{aligned}$$

26. (a) The equation

$$\|f\| = |f(a)| + |f(b)|$$

does not define a norm on $C[a,b]$. For example, the function $f(x) = x^2 - x$ in $C[0,1]$ has the property

$$\|f\| = |f(0)| + |f(1)| = 0$$

however, f is not the zero function.

(b) The expression

$$\|f\| = \int_a^b |f(x)|\, dx$$

defines a norm on $C[a,b]$. To see this we must show that the three conditions in the definition of norm are satisfied.

(i) $\int_a^b |f(x)|\, dx \geq 0$. Equality can occur if and only if f is the zero function. Indeed, if $f(x_0) \neq 0$ for some x_0 in $[a,b]$, then the continuity of $f(x)$ implies that $|f(x)| > 0$ for all x in some interval containing x_0 and consequently $\int_a^b |f(x)|\, dx > 0$.

(ii)

$$\|\alpha f\| = \int_a^b |\alpha f(x)|\, dx = |\alpha| \int_a^b |f(x)|\, dx = |\alpha| \|f\|$$

(iii)

$$\begin{aligned} \|f+g\| &= \int_a^b |f(x) + g(x)|\, dx \\ &\leq \int_a^b (|f(x)| + |g(x)|)\, dx \\ &= \int_a^b |f(x)|\, dx + \int_a^b |g(x)|\, dx \\ &= \|f\| + \|g\| \end{aligned}$$

(c) The expression

$$\|f\| = \max_{a \leq x \leq b} |f(x)|$$

defines a norm on $C[a,b]$. To see this we must verify that three conditions are satisfied.

(i) Clearly $\max_{a \leq x \leq b} |f(x)| \geq 0$. Equality can occur only if f is the zero function.

(ii)

$$\|\alpha f\| = \max_{a \leq x \leq b} |\alpha f(x)| = |\alpha| \max_{a \leq x \leq b} |f(x)| = |\alpha|\, \|f\|$$

(iii)

$$\begin{aligned} \|f+g\| &= \max_{a \leq x \leq b} |f(x) + g(x)| \\ &\leq \max_{a \leq x \leq b} (|f(x)| + |g(x)|) \end{aligned}$$

$$\le \max_{a\le x\le b} |f(x)| + \max_{a\le x\le b} |g(x)|$$
$$= \|f\| + \|g\|$$

27. (a) If $\mathbf{x} \in R^n$, then

$$|x_i| \le \max_{1\le j\le n} |x_j| = \|\mathbf{x}\|_\infty$$

and hence

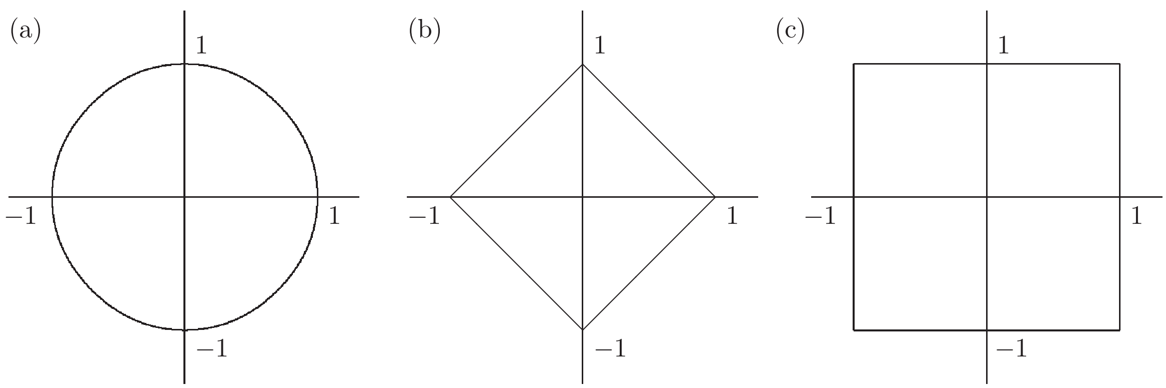
$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i| \le n\|\mathbf{x}\|_\infty$$

(b) $$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \le \left(\sum_{i=1}^{n}(\max_{1\le j\le n} |x_j|)^2\right)^{1/2}$$
$$= (n(\max_{1\le j\le n} |x_j|^2))^{1/2} = \sqrt{n}\|\mathbf{x}\|_\infty$$

If $\mathbf{x}$ is a vector whose entries are all equal to 1 then for this vector equality will hold in parts (a) and (b) since

$$\|\mathbf{x}\|_\infty = 1, \quad \|\mathbf{x}\|_1 = n, \quad \|\mathbf{x}\|_2 = \sqrt{n}$$

28. Each norm produces a different unit "circle".



29. (a) $\langle A\mathbf{x}, \mathbf{y}\rangle = (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^T A^T\mathbf{y} = \langle \mathbf{x}, A^T\mathbf{y}\rangle$
(b) $\langle A^TA\mathbf{x}, \mathbf{x}\rangle = \langle \mathbf{x}, A^TA\mathbf{x}\rangle = \mathbf{x}^T A^TA\mathbf{x} = (A\mathbf{x})^T A\mathbf{x} = \langle A\mathbf{x}, A\mathbf{x}\rangle = \|A\mathbf{x}\|^2$

SECTION 5

2. (a) $\mathbf{u}_1^T\mathbf{u}_1 = \frac{1}{18} + \frac{1}{18} + \frac{16}{18} = 1$

$\mathbf{u}_2^T\mathbf{u}_2 = \frac{4}{9} + \frac{4}{9} + \frac{1}{9} = 1$

$$\mathbf{u}_3^T\mathbf{u}_3 = \frac{1}{2} + \frac{1}{2} + 0 = 1$$

$$\mathbf{u}_1^T\mathbf{u}_2 = \frac{\sqrt{2}}{9} + \frac{\sqrt{2}}{9} - \frac{2\sqrt{2}}{9} = 0$$

$$\mathbf{u}_1^T\mathbf{u}_3 = \frac{1}{6} - \frac{1}{6} + 0 = 0$$

$$\mathbf{u}_2^T\mathbf{u}_3 = \frac{\sqrt{2}}{3} - \frac{\sqrt{2}}{3} + 0 = 0$$

4. (a) $\mathbf{x}_1^T\mathbf{x}_1 = \cos^2\theta + \sin^2\theta = 1$

$\mathbf{x}_2^T\mathbf{x}_2 = (-\sin\theta)^2 + \cos^2\theta = 1$

$\mathbf{x}_1^T\mathbf{x}_2 = -\cos\theta\sin\theta + \sin\theta\cos\theta = 0$

(c)
$$\begin{aligned} c_1^2 + c_2^2 &= (y_1\cos\theta + y_2\sin\theta)^2 + (-y_1\sin\theta + y_2\cos\theta)^2 \\ &= y_1^2\cos^2\theta + 2y_1y_2\sin\theta\cos\theta + y_2^2\sin^2\theta \\ &\quad + y_1^2\sin^2\theta - 2y_1y_2\sin\theta\cos\theta + y_2^2\cos^2\theta \\ &= y_1^2 + y_2^2. \end{aligned}$$

5. If $c_1 = \mathbf{u}^T\mathbf{u}_1 = \frac{1}{2}$ and $c_2 = \mathbf{u}^T\mathbf{u}_2$, then by Theorem 5.5.2

$$\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2$$

It follows from Parseval's formula that

$$1 = \|\mathbf{u}\|^2 = c_1^2 + c_2^2 = \frac{1}{4} + c_2^2$$

Hence

$$|\mathbf{u}^T\mathbf{u}_2| = |c_2| = \frac{\sqrt{3}}{2}$$

. 7 By Parseval's formula

$$c_1^2 + c_2^2 + c_3^2 = \|\mathbf{x}\|^2 = 25$$

It follows from Theorem 5.5.2 that

$$c_1 = \langle \mathbf{u}_1, \mathbf{x}\rangle = 4 \qquad \text{and} \qquad c_2 = \langle \mathbf{u}_2, \mathbf{x}\rangle = 0$$

Plugging these values into Parseval's formula we get

$$16 + 0 + c_3^2 = 25$$

and hence $c_3 = \pm 3$.

8. Since $\{\sin x, \cos x\}$ is an orthonormal set it follows that

$$\langle f, g\rangle = 3\cdot 1 + 2\cdot(-1) = 1$$

9. (a)
$$\begin{aligned} \sin^4 x &= \left(\frac{1-\cos 2x}{2}\right)^2 \\ &= \frac{1}{4}\cos^2 2x - \frac{1}{2}\cos 2x + \frac{1}{4} \end{aligned}$$

$$= \frac{1}{4}\left(\frac{1+\cos 4x}{2}\right) - \frac{1}{2}\cos 2x + \frac{1}{4}$$

$$= \frac{1}{8}\cos 4x - \frac{1}{2}\cos 2x + \frac{3\sqrt{2}}{8}\frac{1}{\sqrt{2}}$$

(b) (i) $\displaystyle\int_{-\pi}^{\pi} \sin^4 x \cos x \, dx = \pi \cdot 0 = 0$

(ii) $\displaystyle\int_{-\pi}^{\pi} \sin^4 x \cos 2x \, dx = \pi(-\frac{1}{2}) = -\frac{\pi}{2}$

(iii) $\displaystyle\int_{-\pi}^{\pi} \sin^4 x \cos 3x \, dx = \pi \cdot 0 = 0$

(iv) $\displaystyle\int_{-\pi}^{\pi} \sin^4 x \cos 4x \, dx = \pi \cdot \frac{1}{8} = \frac{\pi}{8}$

10. The key to seeing why F_8P_8 can be partitioned into block form

$$\begin{pmatrix} F_4 & D_4F_4 \\ F_4 & -D_4F_4 \end{pmatrix}$$

is to note that

$$\omega_8^{2k} = e^{-\frac{4k\pi i}{8}} = e^{-\frac{2k\pi i}{4}} = \omega_4^k$$

and there are repeating patterns in the powers of ω_8. Since

$$\omega_8^4 = -1 \qquad \text{and} \qquad \omega_8^{8n} = e^{-2n\pi i} = 1$$

it follows that

$$\omega_8^{j+4} = -\omega_8^j \qquad \text{and} \qquad \omega_8^{8n+j} = \omega_8^j$$

Using these results let us examine the odd and even columns of F_8. Let us denote the jth column vector of the $m \times m$ Fourier matrix by $\mathbf{f}_j^{(m)}$. The odd columns of the 8×8 Fourier matrix are of the form

$$\mathbf{f}_{2n+1}^{(8)} = \begin{pmatrix} \omega_8^0 \\ \omega_8^{2n} \\ \omega_8^{4n} \\ \omega_8^{6n} \\ \omega_8^{8n} \\ \omega_8^{10n} \\ \omega_8^{12n} \\ \omega_8^{14n} \end{pmatrix} = \begin{pmatrix} 1 \\ \omega_8^{2n} \\ \omega_8^{4n} \\ \omega_8^{6n} \\ 1 \\ \omega_8^{2n} \\ \omega_8^{4n} \\ \omega_8^{6n} \end{pmatrix} = \begin{pmatrix} 1 \\ \omega_4^{n} \\ \omega_4^{2n} \\ \omega_4^{3n} \\ 1 \\ \omega_4^{n} \\ \omega_4^{2n} \\ \omega_4^{3n} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_{n+1}^{(4)} \\ \mathbf{f}_{n+1}^{(4)} \end{pmatrix}$$

for $n = 0, 1, 2, 3$. The even columns are of the form

$$\mathbf{f}^{(8)}_{2n+2} = \begin{pmatrix} \omega_8^0 \\ \omega_8^{2n+1} \\ \omega_8^{2(2n+1)} \\ \omega_8^{3(2n+1)} \\ \omega_8^{4(2n+1)} \\ \omega_8^{5(2n+1)} \\ \omega_8^{6(2n+1)} \\ \omega_8^{7(2n+1)} \end{pmatrix} = \begin{pmatrix} 1 \\ \omega_8\omega_8^{2n} \\ \omega_8^2\omega_8^{4n} \\ \omega_8^3\omega_8^{6n} \\ -1 \\ -\omega_8\omega_8^{2n} \\ -\omega_8^2\omega_8^{4n} \\ -\omega_8^3\omega_8^{6n} \end{pmatrix} = \begin{pmatrix} 1 \\ \omega_8\omega_4^{n} \\ \omega_8^2\omega_4^{2n} \\ \omega_8^3\omega_4^{3n} \\ -1 \\ -\omega_8\omega_4^{n} \\ -\omega_8^2\omega_4^{2n} \\ -\omega_8^3\omega_4^{3n} \end{pmatrix} = \begin{pmatrix} D_4\mathbf{f}^{(4)}_{n+1} \\ -D_4\mathbf{f}^{(4)}_{n+1} \end{pmatrix}$$

for $n = 0, 1, 2, 3$.

11. If Q is orthogonal then

$$(Q^T)^T(Q^T) = QQ^T = QQ^{-1} = I$$

Therefore Q^T is orthogonal.

12. Let θ denote the angle between $\mathbf{x}$ and $\mathbf{y}$ and let θ_1 denote the angle between $Q\mathbf{x}$ and $Q\mathbf{y}$. It follows that

$$\cos\theta_1 = \frac{(Q\mathbf{x})^TQ\mathbf{y}}{\|Q\mathbf{x}\|\|Q\mathbf{y}\|} = \frac{\mathbf{x}^TQ^TQ\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \cos\theta$$

and hence the angles are the same.

13. (a) Use mathematical induction to prove

$$(Q^m)^{-1} = (Q^T)^m = (Q^m)^T, \qquad m = 1, 2, \ldots$$

Proof: The case $m = 1$ follows from Theorem 5.5.5. If for some positive integer k

$$(Q^k)^{-1} = (Q^T)^k = (Q^k)^T$$

then

$$(Q^T)^{k+1} = Q^T(Q^T)^k = Q^T(Q^k)^T = (Q^kQ)^T = (Q^{k+1})^T$$

and

$$(Q^T)^{k+1} = Q^T(Q^T)^k = Q^{-1}(Q^k)^{-1} = (Q^kQ)^{-1} = (Q^{k+1})^{-1}$$

(b) Prove: $\|Q^m\mathbf{x}\| = \|\mathbf{x}\|$ for $m = 1, 2, \ldots$.
Proof: In the case $m = 1$

$$\|Q\mathbf{x}\|^2 = (Q\mathbf{x})^TQ\mathbf{x} = \mathbf{x}^TQ^TQ\mathbf{x} = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2$$

and hence

$$\|Q\mathbf{x}\| = \|\mathbf{x}\|$$

If $\|Q^k\mathbf{y}\| = \|\mathbf{y}\|$ for any $\mathbf{y} \in R^n$, then in particular, if $\mathbf{x}$ is an arbitrary vector in R^n and we define $\mathbf{y} = Q\mathbf{x}$, then

$$\|Q^{k+1}\mathbf{x}\| = \|Q^k(Q\mathbf{x})\| = \|Q^k\mathbf{y}\| = \|\mathbf{y}\| = \|Q\mathbf{x}\| = \|\mathbf{x}\|$$

14. $H^T = (I - 2\mathbf{u}\mathbf{u}^T)^T = I^T - 2(\mathbf{u}^T)^T\mathbf{u}^T = I - 2\mathbf{u}\mathbf{u}^T = H$

$$\begin{aligned} H^T H = H^2 & \\ &= (I - 2\mathbf{u}\mathbf{u}^T)^2 \\ &= I - 4\mathbf{u}\mathbf{u}^T + 4\mathbf{u}\mathbf{u}^T\mathbf{u}\mathbf{u}^T \\ &= I - 4\mathbf{u}\mathbf{u}^T + 4\mathbf{u}\mathbf{u}^T \\ &= I \end{aligned}$$

15. Since $Q^TQ = I$, it follows that

$$[\det(Q)]^2 = \det(Q^T)\det(Q) = \det(I) = 1$$

Thus $\det(Q) = \pm 1$.

16. (a) Let Q_1 and Q_2 be orthogonal $n \times n$ matrices and let $Q = Q_1Q_2$. It follows that

$$Q^TQ = (Q_1Q_2)^TQ_1Q_2 = Q_2^TQ_1^TQ_1Q_2 = I$$

Therefore Q is orthogonal.

(b) Yes. Let P_1 and P_2 be permutation matrices. The columns of P_1 are the same as the columns of I, but in a different order. Postmultiplication of P_1 by P_2 reorders the columns of P_1. Thus P_1P_2 is a matrix formed by reordering the columns of I and hence is a permutation matrix.

17. There are $n!$ permutations of any set with n distinct elements. Therefore there are $n!$ possible permutations of the row vectors of the $n \times n$ identity matrix and hence the number of $n \times n$ permutation matrices is $n!$.

18. A permutation P is an orthogonal matrix so $P^T = P^{-1}$ and if P is a symmetric permutation matrix then $P = P^T = P^{-1}$ and hence

$$P^2 = P^TP = P^{-1}P = I$$

So for a symmetric permutation matrix we have

$$P^{2k} = (P^2)^k = I^k = I \qquad \text{and} \qquad P^{2k+1} = PP^{2k} = PI = P$$

19.

$$\begin{aligned} I = UU^T &= (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)\begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{pmatrix} \\ &= \mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T + \cdots + \mathbf{u}_n\mathbf{u}_n^T \end{aligned}$$

20. The proof is by induction on n. If $n = 1$, then Q must be either (1) or (-1). Assume the result holds for all $k \times k$ upper triangular orthogonal matrices and let Q be a $(k+1)\times(k+1)$ matrix that is upper triangular and orthogonal. Since Q is upper triangular its first column must be a multiple of $\mathbf{e}_1$. But Q

is also orthogonal, so $\mathbf{q}_1$ is a unit vector. Thus $\mathbf{q}_1 = \pm\mathbf{e}_1$. Furthermore, for $j = 2, \ldots, n$

$$q_{1j} = \mathbf{e}_1^T \mathbf{q}_j = \pm\mathbf{q}_1^T \mathbf{q}_j = 0$$

Thus Q must be of the form

$$Q = \begin{pmatrix} \pm 1 & 0 & 0 & \cdots & 0 \\ \mathbf{0} & \mathbf{p}_2 & \mathbf{p}_3 & \cdots & \mathbf{p}_{k+1} \end{pmatrix}$$

The matrix $P = (\mathbf{p}_2, \mathbf{p}_3, \ldots, \mathbf{p}_{k+1})$ is a $k \times k$ matrix that is both upper triangular and orthogonal. By the induction hypothesis P must be a diagonal matrix with diagonal entries equal to ± 1. Thus Q must also be a diagonal matrix with ± 1's on the diagonal.

21. (a) The columns of A form an orthonormal set since

$$\begin{aligned} \mathbf{a}_1^T \mathbf{a}_2 &= -\frac{1}{4} - \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 0 \\ \mathbf{a}_1^T \mathbf{a}_1 &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1 \\ \mathbf{a}_2^T \mathbf{a}_2 &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1 \end{aligned}$$

22. (b)

(i) $A\mathbf{x} = P\mathbf{b} = (2, 2, 0, 0)^T$

(ii) $A\mathbf{x} = P\mathbf{b} = \left(\frac{3}{2}, \frac{3}{2}, \frac{7}{2}, \frac{7}{2}\right)^T$

(iii) $A\mathbf{x} = P\mathbf{b} = (1, 1, 2, 2)^T$

23. (a) One can find a basis for $N(A^T)$ in the usual way by computing the reduced row echelon form of A^T.

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Setting the free variables equal to one and solving for the lead variables, we end up with basis vectors $\mathbf{x}_1 = (-1, 1, 0, 0)^T$, $\mathbf{x} = (0, 0, -1, 1)^T$. Since these vectors are already orthogonal we need only normalize to obtain an orthonormal basis for $N(A^T)$.

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}(-1, 1, 0, 0)^T \qquad \mathbf{u}_2 = \frac{1}{\sqrt{2}}(0, 0, -1, 1)^T$$

24. (a) Let U_1 be a matrix whose columns form an orthonormal basis for $R(A)$ and let U_2 be a matrix whose columns form an orthonormal basis for $N(A^T)$. If we set $U = (U_1, U_2)$, then since $R(A)$ and $N(A^T)$ are orthogonal complements in R^n, it follows that U is an orthogonal matrix. The unique projection matrix P onto $R(A)$ is given $P = U_1 U_1^T$ and the projection matrix onto $N(A^T)$ is given by $U_2 U_2^T$. Since U is orthogonal it follows that

$$I = UU^T = U_1U_1^T + U_2U_2^T = P + U_2U_2^T$$

Thus the projection matrix onto $N(A^T)$ is given by

$$U_2U_2^T = I - P$$

(b) The proof here is essentially the same as in part (a). Let V_1 be a matrix whose columns form an orthonormal basis for $R(A^T)$ and let V_2 be a matrix whose columns form an orthonormal basis for $N(A)$. If we set $V = (V_1, V_2)$, then since $R(A^T)$ and $N(A)$ are orthogonal complements in R^m, it follows that V is an orthogonal matrix. The unique projection matrix Q onto $R(A^T)$ is given $Q = V_1V_1^T$ and the projection matrix onto $N(A)$ is given by $V_2V_2^T$. Since V is orthogonal it follows that

$$I = VV^T = V_1V_1^T + V_2V_2^T = Q + V_2V_2^T$$

Thus the projection matrix onto $N(A)$ is given by

$$V_2V_2^T = I - Q$$

25. (a) If U is a matrix whose columns form an orthonormal basis for S, then the projection matrix P corresponding to S is given by $P = UU^T$. It follow that

$$P^2 = (UU^T)(UU^T) = U(U^TU)U^T = UIU^T = P$$

(b) $P^T = (UU^T)^T = (U^T)^TU^T = UU^T = P$

26. The (i, j) entry of A^TA will be $\mathbf{a}_i^T\mathbf{a}_j$. This will be 0 if $i \neq j$. Thus A^TA is a diagonal matrix with diagonal elements $\mathbf{a}_1^T\mathbf{a}_1, \mathbf{a}_2^T\mathbf{a}_2, \ldots, \mathbf{a}_n^T\mathbf{a}_n$. The ith entry of $A^T\mathbf{b}$ is $\mathbf{a}_i^T\mathbf{b}$. Thus if $\hat{\mathbf{x}}$ is the solution to the normal equations, its ith entry will be

$$\hat{\mathbf{x}}_i = \frac{\mathbf{a}_i^T\mathbf{b}}{\mathbf{a}_i^T\mathbf{a}_i} = \frac{\mathbf{b}^T\mathbf{a}_i}{\mathbf{a}_i^T\mathbf{a}_i}$$

27. (a) $\langle 1, x\rangle = \int_{-1}^{1} 1x\,dx = \left.\frac{x^2}{2}\right|_{-1}^{1} = 0$

28. (a) $\langle 1, 2x-1\rangle = \int_0^1 1\cdot(2x-1)dx = \left. x^2 - x\right|_0^1 = 0$

(b) $\|1\|^2 = \langle 1, 1\rangle = \int_0^1 1\cdot 1\,dx = \left. x\right|_0^1 = 1$

$\|2x-1\|^2 = \int_0^1 (2x-1)^2dx = \frac{1}{3}$

Therefore

$$\|1\| = 1 \qquad \text{and} \qquad \|2x-1\| = \frac{1}{\sqrt{3}}$$

(c) The best least squares approximation to $\sqrt{x}$ from S is given by

$$\ell(x) = c_1 1 + c_2\sqrt{3}(2x-1)$$

where

$$\begin{aligned} c_1 &= \langle 1, x^{1/2} \rangle = \int_0^1 1\, x^{1/2} dx = \frac{2}{3} \\ c_2 &= \langle \sqrt{3}(2x-1), x^{1/2} \rangle = \int_0^1 \sqrt{3}(2x-1) x^{1/2} dx = \frac{2\sqrt{3}}{15} \end{aligned}$$

Thus

$$\begin{aligned} \ell(x) &= \frac{2}{3} \cdot 1 + \frac{2\sqrt{3}}{15}(\sqrt{3}(2x-1)) \\ &= \frac{4}{5}(x + \frac{1}{3}) \end{aligned}$$

29. We saw in Example 3 that $\{1/\sqrt{2}, \cos x, \cos 2x, \ldots, \cos nx\}$ is an orthonormal set. In Section 4, Exercise 9 we saw that the functions $\cos kx$ and $\sin jx$ were orthogonal unit vectors in $C[-\pi, \pi]$. Furthermore

$$\left\langle \frac{1}{\sqrt{2}}, \sin jx \right\rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{1}{\sqrt{2}} \sin jx \; dx = 0$$

Therefore $\{1/\sqrt{2}, \cos x, \cos 2x, \ldots, \cos nx, \sin x, \sin 2x, \ldots, \sin nx\}$ is an orthonormal set of vectors.

30. The coefficients of the best approximation are given by

$$a_0 = \langle 1, |x| \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} 1 \cdot |x| \, dx = \frac{2}{\pi} \int_0^{\pi} x \, dx = \pi$$

$$a_1 = \langle \cos x, |x| \rangle = \frac{2}{\pi} \int_0^{\pi} x \cos x \, dx = -\frac{4}{\pi}$$

$$a_2 = \frac{2}{\pi} \int_0^{\pi} x \cos 2x \, dx = 0$$

To compute the coefficients of the sin terms we must integrate $x \sin x$ and $x \sin 2x$ from $-\pi$ to π. Since both of these are odd functions the integrals will be 0. Therefore $b_1 = b_2 = 0$. The best trigonometric approximation of degree 2 or less is given by

$$p(x) = \frac{\pi}{2} - \frac{4}{\pi} \cos x$$

31. If $\mathbf{u} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k$ is an element of S_1 and $\mathbf{v} = c_{k+1}\mathbf{x}_{k+1} + c_{k+2}\mathbf{x}_{k+2} + \cdots + c_n\mathbf{x}_n$ is an element of S_2, then

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} \rangle &= \left\langle \sum_{i=1}^{k} c_i\mathbf{x}_i, \sum_{j=k+1}^{n} c_j\mathbf{x}_j \right\rangle \\ &= \sum_{k=1}^{k} \sum_{j=k+1}^{n} c_i c_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ &= 0 \end{aligned}$$

32. (a) By Theorem 5.5.2,

$$\begin{aligned}\mathbf{x} &= \sum_{i=1}^{n}\langle \mathbf{x}, \mathbf{x}_i\rangle \mathbf{x}_i \\ &= \sum_{i=1}^{k}\langle \mathbf{x}, \mathbf{x}_i\rangle \mathbf{x}_i + \sum_{i=k+1}^{n}\langle \mathbf{x}, \mathbf{x}_i\rangle \mathbf{x}_i \\ &= \mathbf{p}_1 + \mathbf{p}_2\end{aligned}$$

(b) It follows from Exercise 31 that $S_2 \subset S_1^{\perp}$. On the other hand if $\mathbf{x} \in S_1^{\perp}$ then by part (a) $\mathbf{x} = \mathbf{p}_1 + \mathbf{p}_2$. Since $\mathbf{x} \in S_1^{\perp}$, $\langle \mathbf{x}, \mathbf{x}_i\rangle = 0$ for $i = 1, \ldots, k$. Thus $\mathbf{p}_1 = \mathbf{0}$ and $\mathbf{x} = \mathbf{p}_2 \in S_2$. Therefore $S_2 = S^{\perp}$.

33. Let

$$\mathbf{u}_i = \frac{1}{\|\mathbf{x}_i\|}\mathbf{x}_i \quad \text{for} \quad i = 1, \ldots, n$$

By Theorem 5.5.8 the best least squares approximation to $\mathbf{x}$ from S is given by

$$\begin{aligned}\mathbf{p} = \sum_{i=1}^{n}\langle \mathbf{x}, \mathbf{u}_i\rangle \mathbf{u}_i &= \sum_{i=1}^{n}\frac{1}{\|\mathbf{x}_i\|^2}\langle \mathbf{x}, \mathbf{x}_i\rangle \mathbf{x}_i \\ &= \sum_{i=1}^{n}\frac{\langle \mathbf{x}, \mathbf{x}_i\rangle}{\langle \mathbf{x}_i, \mathbf{x}_i\rangle}\mathbf{x}_i.\end{aligned}$$

SECTION 6

9. $r_{11} = \|\mathbf{x}_1\| = 5$

$\mathbf{q}_1 = \dfrac{1}{r_{11}}\mathbf{x}_1 = \left(\dfrac{4}{5}, \dfrac{2}{5}, \dfrac{2}{5}, \dfrac{1}{5}\right)^T$

$r_{12} = \mathbf{q}_1^T\mathbf{x}_2 = 2 \quad \text{and} \quad r_{13} = \mathbf{q}_1^T\mathbf{x}_3 = 1$

$\mathbf{x}_2^{(1)} = \mathbf{x}_2 - r_{12}\mathbf{q}_1 = \left(\dfrac{2}{5}, -\dfrac{4}{5}, -\dfrac{4}{5}, \dfrac{8}{5}\right)^T, \mathbf{x}_3^{(1)} = \mathbf{x}_3 - r_{13}\mathbf{q}_1 = \left(\dfrac{1}{5}, \dfrac{3}{5}, -\dfrac{7}{5}, \dfrac{4}{5}\right)^T$

$r_{22} = \|\mathbf{x}_2^{(1)}\| = 2$

$\mathbf{q}_2 = \dfrac{1}{r_{22}}\mathbf{x}_2^{(1)} = \left(\dfrac{1}{5}, -\dfrac{2}{5}, -\dfrac{2}{5}, \dfrac{4}{5}\right)^T$

$r_{23} = \mathbf{x}_3^T\mathbf{q}_2 = 1$

$\mathbf{x}_3^{(2)} = \mathbf{x}_3^{(1)} - r_{23}\mathbf{q}_2 = (0,\ 1,\ -1,\ 0)^T$

$r_{33} = \|\mathbf{x}_3^{(2)}\| = \sqrt{2}$

$\mathbf{q}_3 = \dfrac{1}{r_{33}}\mathbf{x}_3^{(2)} = \left(0, \dfrac{1}{\sqrt{2}}, -\dfrac{1}{\sqrt{2}}, 0\right)^T$

10. Given a basis $\{x_1, \ldots, x_n\}$, one can construct an orthonormal basis using either the classical Gram–Schmidt process or the modified process. When carried out in exact arithmetic both methods will produce the same orthonormal set $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$.

Proof: The proof is by induction on n. In the case $n = 1$, the vector $\mathbf{q}_1$ is computed in the same way for both methods.

$$\mathbf{q}_1 = \frac{1}{r_{11}}\mathbf{x}_1 \quad \text{where} \quad r_{11} = \|\mathbf{x}\|_1$$

Assume $\mathbf{q}_1, \ldots, \mathbf{q}_k$ are the same for both methods. In the classical Gram–Schmidt process one computes $\mathbf{q}_{k+1}$ as follows: Set

$$\begin{aligned}
r_{i,k+1} &= \langle \mathbf{x}_{k+1}, \mathbf{q}_i \rangle, \qquad i = 1, \ldots, k \\
\mathbf{p}_k &= r_{1,k+1}\mathbf{q}_1 + r_{2,k+1}\mathbf{q}_2 + \cdots + r_{k,k+1}\mathbf{q}_k \\
r_{k+1,k+1} &= \|\mathbf{x}_{k+1} - \mathbf{p}_k\| \\
\mathbf{q}_{k+1} &= \frac{1}{r_{k+1,k+1}}(\mathbf{x}_{k+1} - \mathbf{p}_k)
\end{aligned}$$

Thus

$$\mathbf{q}_{k+1} = \frac{1}{r_{k+1,k+1}}(\mathbf{x}_{k+1} - r_{1,k+1}\mathbf{q}_1 - r_{2,k+1}\mathbf{q}_2 - \cdots - r_{k,k+1}\mathbf{q}_k)$$

In the modified version, at step 1 the vector $r_{1,k+1}\mathbf{q}_1$ is subtracted from $\mathbf{x}_{k+1}$.

$$\mathbf{x}_{k+1}^{(1)} = \mathbf{x}_{k+1} - r_{1,k+1}\mathbf{q}_1$$

At the next step $r_{2,k+1}\mathbf{q}_2$ is subtracted from $\mathbf{x}_{k+1}^{(1)}$.

$$\begin{aligned}
\mathbf{x}_{k+1}^{(2)} &= \mathbf{x}_{k+1}^{(1)} - r_{2,k+1}\mathbf{q}_2 \\
&= \mathbf{x}_{k+1} - r_{1,k+1}\mathbf{q}_1 - r_{2,k+1}\mathbf{q}_2
\end{aligned}$$

In general after k steps we have

$$\begin{aligned}
\mathbf{x}_{k+1}^{(k)} &= \mathbf{x}_{k+1} - r_{1,k+1}\mathbf{q}_1 - r_{2,k+1}\mathbf{q}_2 - \cdots - r_{k,k+1}\mathbf{q}_k \\
&= \mathbf{x}_{k+1} - \mathbf{p}_k
\end{aligned}$$

In the last step we set

$$r_{k+1,k+1} = \|\mathbf{x}_{k+1}^{(k)}\| = \|\mathbf{x}_{k+1} - \mathbf{p}_k\|$$

and set

$$\mathbf{q}_{k+1} = \frac{1}{r_{k+1,k+1}}\mathbf{x}_{k+1}^{(k)} = \frac{1}{r_{k+1,k+1}}(\mathbf{x}_{k+1} - \mathbf{p}_k)$$

Thus $\mathbf{q}_{k+1}$ is the same as in the classical Gram–Schmidt process.

11. If the Gram-Schmidt process is applied to a set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\mathbf{v}_3$ is in $\text{Span}(\mathbf{v}_1, \mathbf{v}_2)$, then the process will break down at the third step. If $\mathbf{u}_1$, $\mathbf{u}_2$ have been constructed so that they form an orthonormal basis for $S_2 = \text{Span}(\mathbf{v}_1, \mathbf{v}_2)$, then the projection $\mathbf{p}_2$ of $\mathbf{v}_3$ onto S_2 is $\mathbf{v}_3$ (since $\mathbf{v}_3$ is already in S_2). Thus $\mathbf{v}_3 - \mathbf{p}_2$ will be the zero vector and hence we cannot normalize to obtain a unit vector $\mathbf{u}_3$.

12. (a) Since

$$\mathbf{p} = c_1\mathbf{q}_1 + c_2\mathbf{q}_2 + \cdots + c_n\mathbf{q}_n$$

is the projection of $\mathbf{b}$ onto $R(A)$ and $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ form an orthonormal basis for $R(A)$, it follows that

$$c_j = \mathbf{q}_j^T\mathbf{b} \qquad j = 1, \ldots, n$$

and hence

$$\mathbf{c} = Q^T\mathbf{b}$$

(b) $\mathbf{p} = c_1\mathbf{q}_1 + c_2\mathbf{q}_2 + \cdots + c_n\mathbf{q}_n = Q\mathbf{c} = QQ^T\mathbf{b}$

(c) Both $A(A^TA)^{-1}A^T$ and QQ^T are projection matrices that project vectors onto $R(A)$. Since the projection matrix is unique for a given subspace it follows that

$$QQ^T = A(A^TA)^{-1}A^T$$

13. **(a)** If $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthonormal basis for V then by Theorem 3.4.4 it can be extended to form a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_{k+1}, \mathbf{u}_{k+2}, \ldots, \mathbf{u}_m\}$ for U. If we apply the Gram-Schmidt process to this basis, then since $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are already orthonormal vectors, they will remain unchanged and we with end up with an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_m\}$.

(b) If $\mathbf{u}$ is any vector in U, we can write

$$(3) \qquad \mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k + c_{k+1}\mathbf{v}_{k+1} + \cdots + c_m\mathbf{v}_m = \mathbf{v} + \mathbf{w}$$

where

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \in V \qquad \text{and} \qquad \mathbf{w} = c_{k+1}\mathbf{v}_{k+1} + \cdots + c_m\mathbf{v}_m) \in W$$

Therefore, $U = V + W$. The representation (3) is unique. Indeed if

$$\mathbf{u} = \mathbf{v} + \mathbf{w} = \mathbf{x} + \mathbf{y}$$

where $\mathbf{v}, \mathbf{x}$ are in V and $\mathbf{w}, \mathbf{y}$ are in W, then

$$\mathbf{v} - \mathbf{x} = \mathbf{y} - \mathbf{w}$$

and hence $\mathbf{v} - \mathbf{x} \in V \cap W$. Since V and W are orthogonal subspaces we have $V \cap W = \{\mathbf{0}\}$ and hence $\mathbf{v} = \mathbf{x}$. By the same reasoning $\mathbf{w} = \mathbf{y}$. It follows then that $U = V \oplus W$.

14. Let $m = \dim U$, $k = \dim V$, and $W = U \cap V$. If $\dim W = r > 0$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is a basis for W, then by Exercise 13(a) we can extend this basis to an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{v}_{r+1}, \ldots, \mathbf{v}_k\}$ for V. Let

$$V_1 = \operatorname{Span}(\mathbf{v}_{r+1}, \ldots, \mathbf{v}_k)$$

By Exercise 13(b) we have $V = W \oplus V_1$. We claim that $U + V = U \oplus V_1$. Since V_1 is a subspace of V it follows that $U + V_1$ is a subspace of $U + V$. On the other hand, if $\mathbf{x}$ is in $U + V$ then

$$\mathbf{x} = \mathbf{u} + \mathbf{v} = \mathbf{u} + (\mathbf{w} + \mathbf{v}_1) = (\mathbf{u} + \mathbf{w}) + \mathbf{v}_1$$

where $\mathbf{u} \in U$, $\mathbf{v} \in V$, $\mathbf{w} \in W$, and $\mathbf{v}_1 \in V_1$. Since $\mathbf{u} + \mathbf{w}$ is in U it follows that $\mathbf{x}$ is in $U + V_1$ and hence $U + V = U + V_1$. To show that we have a direct sum we must show that $U \cap V_1 = \{\mathbf{0}\}$. If $\mathbf{z} \in U \cap V_1$ then $\mathbf{z}$ is also in the larger subspace $W = U \cap V$. So $\mathbf{z}$ is in both V_1 and W. However, by construction V_1 is orthogonal to W, so the intersection of the two subspaces must be $\{\mathbf{0}\}$. Therefore $U \cap V_1 = \{\mathbf{0}\}$. It follows then that

$$U + V = U \oplus V$$

and hence

$$\begin{aligned} \dim(U+V) &= \dim(U \oplus V) = \dim U + \dim V_1 \\ &= m + (k-r) = m + k - r \\ &= \dim U + \dim V - \dim(U \cap V) \end{aligned}$$

SECTION 7

3. Let $x = \cos\theta$.

(a) $$\begin{aligned} 2T_m(x)T_n(x) &= 2\cos m\theta \cos n\theta \\ &= \cos(m+n)\theta + \cos(m-n)\theta \\ &= T_{m+n}(x) + T_{m-n}(x) \end{aligned}$$

(b) $T_m(T_n(x)) = T_m(\cos n\theta) = \cos(mn\theta) = T_{mn}(x)$

5. $p_n(x) = a_n x^n + q(x)$ where degree $q(x) < n$. By Theorem 5.7.1, $\langle q, p_n \rangle = 0$. It follows then that

$$\begin{aligned} \|p_n\|^2 &= \langle a_n x^n + q(x), p(x) \rangle \\ &= a_n \langle x^n, p_n \rangle + \langle q, p_n \rangle \\ &= a_n \langle x^n, p_n \rangle \end{aligned}$$

6. (b) $$\begin{aligned} U_{n-1}(x) &= \frac{1}{n} T_n'(x) \\ &= \frac{1}{n} \frac{dT_n}{d\theta} \Big/ \frac{dx}{d\theta} \\ &= \frac{\sin n\theta}{\sin\theta} \end{aligned}$$

7. (a) $$\begin{aligned} U_n(x) - xU_{n-1}(x) &= \frac{\sin(n+1)\theta}{\sin\theta} - \frac{\cos\theta \sin n\theta}{\sin\theta} \\ &= \frac{\sin n\theta \cos\theta + \cos n\theta \sin\theta - \cos\theta \sin n\theta}{\sin\theta} \\ &= \cos n\theta \\ &= T_n(x) \end{aligned}$$

(b) $$\begin{aligned} U_n(x) + U_{n-2}(x) &= \frac{\sin(n+1)\theta + \sin(n-1)\theta}{\sin\theta} \\ &= \frac{2\sin n\theta \cos\theta}{\sin\theta} \\ &= 2xU_{n-1}(x) \end{aligned}$$

$$U_n(x) = 2xU_{n-1}(x) - U_{n-2}(x)$$

8. $$\begin{aligned} \langle U_n, U_m\rangle &= \int_{-1}^{1} U_n(x)U_m(x)(1-x^2)^{1/2}dx \\ &= \int_0^{\pi} \sin[(n+1)\theta]\sin[(m+1)\theta]d\theta \quad (x=\cos\theta) \\ &= 0 \quad \text{if} \quad m \neq n \end{aligned}$$

9. (i) $n=0$, $y=1$, $y'=0$, $y''=0$
$(1-x^2)y'' - 2xy' + 0\cdot 1\cdot 1 = 0$

(ii) $n=1$, $y=P_1(x)=x$, $y'=1$, $y''=0$
$(1-x^2)\cdot 0 - 2x\cdot 1 + 1\cdot 2x = 0$

(iii) $n=2$, $y=P_2(x)=\dfrac{3}{2}\left(x^2-\dfrac{1}{3}\right)$, $y'=3x$, $y''=3$
$(1-x^2)\cdot 3 - 2x\cdot 3x + 6\cdot\dfrac{3}{2}\left(x^2-\dfrac{1}{3}\right) = 0$

10. (a) Prove: $H_n'(x) = 2nH_{n-1}(x)$, $n=0,1,2,\ldots$.
Proof: The proof is by mathematical induction. In the case $n=0$

$$H_0'(x) = 0 = 2nH_{-1}(x)$$

Assume

$$H_k'(x) = 2kH_{k-1}(x)$$

for all $k \leq n$.

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$$

Differentiating both sides we get

$$\begin{aligned} H_{n+1}'(x) &= 2H_n + 2xH_n' - 2nH_{n-1}' \\ &= 2H_n + 2x[2nH_{n-1}] - 2n[2(n-1)H_{n-2}] \\ &= 2H_n + 2n[2xH_{n-1} - 2(n-1)H_{n-2}] \\ &= 2H_n + 2nH_n \\ &= 2(n+1)H_n \end{aligned}$$

(b) Prove: $H_n''(x) - 2xH_n'(x) + 2nH_n(x) = 0$, $n=0,1,\ldots$.
Proof: It follows from part (a) that

$$\begin{aligned} H_n'(x) &= 2nH_{n-1}(x) \\ H_n''(x) &= 2nH_{n-1}'(x) = 4n(n-1)H_{n-2}(x) \end{aligned}$$

Therefore

$$\begin{aligned} &H_n''(x) - 2xH_n'(x) + 2nH_n(x) \\ &\quad= 4n(n-1)H_{n-2}(x) - 4xnH_{n-1}(x) + 2nH_n(x) \\ &\quad= 2n[H_n(x) - 2xH_{n-1}(x) + 2(n-1)H_{n-2}(x)] \\ &\quad= 0 \end{aligned}$$

12. If $f(x)$ is a polynomial of degree less than n and $P(x)$ is the Lagrange interpolating polynomial that agrees with $f(x)$ at $x_1, \ldots, x_n$, then degree $P(x) \le n-1$. If we set

$$h(x) = P(x) - f(x)$$

then the degree of h is also $\le n-1$ and

$$h(x_i) = P(x_i) - f(x_i) = 0 \qquad i = 1, \ldots, n$$

Therefore h must be the zero polynomial and hence

$$P(x) = f(x)$$

15. (a) The quadrature formula approximates the integral of $f(x)$ by a sum which is equal to the exact value of the integral of Lagrange polynomial that interpolates f at the given points. In the case where f is a polynomial of degree less than n, the Lagrange polynomial will be equal to f, so the quadrature formula will yield the exact answer.

(b) If we take the constant function $f(x) = 1$ and apply the quadrature formula we get

$$\begin{aligned} \int_{-1}^{1} f(x)dx &= A_1 f(x_1) + A_2 f(x_2) + \cdots + A_n f(x_n) \\ \int_{-1}^{1} 1dx &= A_1 \cdot 1 + A_2 \cdot 1 + \cdots + A_n \cdot 1 \\ 2 &= A_1 + A_2 + \cdot + A_n \end{aligned}$$

16. (a) If $j \ge 1$ then the Legendre polynomial P_j is orthogonal to $P_0 = 1$. Thus we have

$$(4) \qquad \int_{-1}^{1} P_j(x)dx = \int_{-1}^{1} P_j(x)P_0(x)dx = \langle P_j, P_0 \rangle = 0 \quad (j \ge 1)$$

The n-point Gauss-Legendre quadrature formula will yield the exact value of the integral of $f(x)$ whenever $f(x)$ is a polynomial of degree less than $2n$. So in particular for $f(x) = P_j(x)$ we have

$$(5) \int_{-1}^{1} P_j(x)dx = P_j(x_1)A_1 + P_j(x_2)A_2 + \cdots + P_j(x_n)A_n \quad (0 \le j < 2n)$$

It follows from (4) and (5) that

$$P_j(x_1)A_1 + P_j(x_2)A_2 + \cdots + P_j(x_n)A_n = 0 \quad \text{for} \;\; 1 \le j < 2n$$

(b)

$$\begin{aligned} A_1 + A_2 + \cdots + A_n &= 2 \\ P_1(x_1)A_1 + P_1(x_2)A_2 + \cdots + P_1(x_n)A_n &= 0 \\ &\vdots \\ P_{n-1}(x_1)A_1 + P_{n-1}(x_2)A_2 + \cdots + P_{n-1}(x_n)A_n &= 0 \end{aligned}$$

17. (a) If $\|Q_j\| = 1$ for each j, then in the recursion relation we will have

$$\gamma_k = \frac{\langle Q_k, Q_k \rangle}{\langle Q_{k-1}, Q_{k-1} \rangle} = 1 \quad (k \geq 1)$$

and hence the recursion relation for the orthonormal sequence simplifies to

$$\alpha_{k+1} Q_{k+1}(x) = (x - \beta_{k+1} Q_k(x) - \alpha_k Q_{k-1}(x) \quad (k \geq 0)$$

where Q_{-1} is taken to be the zero polynomial.

(b) For $k = 0, \ldots, n-1$ we can rewrite the recursion relation in part (a) in the form

$$\alpha_k Q_{k-1}(x) + \beta_{k+1} Q_k(x) + \alpha_{k+1} Q_{k+1}(x) = x Q_k(x)$$

Let λ be any root of Q_n and let us plug it into each of the n-equations. Note that the first equation ($k = 0$) will be

$$\beta_1 Q_0(\lambda) + \alpha_1 Q_1(\lambda) = \lambda Q_0(\lambda)$$

since Q_{-1} is the zero polynomial. For $(2 \leq k \leq n-2)$ intermediate equations are all of the form

$$\alpha_k Q_{k-1}(\lambda) + \beta_{k+1} Q_k(\lambda) + \alpha_{k+1} Q_{k+1}(\lambda) = \lambda Q_k(\lambda)$$

The last equation ($k = n-1$) will be

$$\alpha_{n-1} Q_{n-2}(\lambda) + \beta_n Q_{n-1}(\lambda) = \lambda Q_{n-1}(\lambda)$$

since $Q_n(\lambda) = 0$. We now have a system of n equations in the variable λ. If we rewrite it in matrix form we get

$$\begin{pmatrix} \beta_1 & \alpha_1 & & & \\ \alpha_1 & \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \alpha_{n-2} & \beta_{n-1} & \alpha_{n-1} \\ & & & \alpha_{n-1} & \beta_n \end{pmatrix} \begin{pmatrix} Q_0(\lambda) \\ Q_1(\lambda) \\ \vdots \\ Q_{n-2}(\lambda) \\ Q_{n-1}(\lambda) \end{pmatrix} = \lambda \begin{pmatrix} Q_0(\lambda) \\ Q_1(\lambda) \\ \vdots \\ Q_{n-2}(\lambda) \\ Q_{n-1}(\lambda) \end{pmatrix}$$

MATLAB EXERCISES

1. (b) By the Cauchy-Schwarz Inequality

$$|\mathbf{x}^T \mathbf{y}| \leq \|\mathbf{x}\| \|\mathbf{y}\|$$

Therefore

$$|t| = \frac{|\mathbf{x}^T \mathbf{y}|}{\|\mathbf{x}\| \, \|\mathbf{y}\|} \leq 1$$

3. (c) From the graph it should be clear that you get a better fit at the bottom of the atmosphere.

5. (a) A is the product of two random matrices. One would expect that both of the random matrices will have full rank, that is, rank 2. Since the row vectors of A are linear combinations of the row vectors of the second random matrix, one would also expect that A would have rank 2. If the rank of A is 2, then the nullity of A should be $5 - 2 = 3$.

(b) Since the column vectors of Q form an orthonormal basis for $R(A)$ and the column vectors of W form an orthonormal basis for $N(A^T) = R(A)^{\perp}$, the column vectors of $S = (Q \;\; W)$ form an orthonormal basis for R^5 and hence S is an orthogonal matrix. Each column vector of W is in $N(A^T)$. thus it follows that

$$A^T W = O$$

and

$$W^T A = (A^T W)^T = O^T$$

(c) Since S is an orthogonal matrix, we have

$$I = SS^T = (Q \;\; W) \begin{pmatrix} Q^T \\ W^T \end{pmatrix} = QQ^T + WW^T$$

Thus

$$QQ^T = I - WW^T$$

and it follows that

$$QQ^T A = A - WW^T A = A - WO = A$$

(d) If $\mathbf{b} \in R(A)$, then $\mathbf{b} = A\mathbf{x}$ for some $\mathbf{x} \in R^5$. It follows from part (c) that

$$QQ^T\mathbf{b} = QQ^T(A\mathbf{x}) = (QQ^T A)\mathbf{x} = A\mathbf{x} = \mathbf{b}$$

Alternatively, one could also argue that since $\mathbf{b} \in N(A^T)^{\perp}$ and the columns of W form an orthonormal basis for $N(A^T)$

$$W^T\mathbf{b} = \mathbf{0}$$

and hence it follows that

$$QQ^T\mathbf{b} = (I - WW^T)\mathbf{b} = \mathbf{b}$$

(e) If $\mathbf{q}$ is the projection of $\mathbf{c}$ onto $R(A)$ and $\mathbf{r} = \mathbf{c} - \mathbf{q}$, then

$$\mathbf{c} = \mathbf{q} + \mathbf{r}$$

and $\mathbf{r}$ is the projection of $\mathbf{c}$ onto $N(A^T)$.

(f) Since the projection of a vector onto a subspace is unique, $\mathbf{w}$ must equal $\mathbf{r}$.

(g) To compute the projection matrix U, set

$$U = Y * Y'$$

Since $\mathbf{y}$ is already in $R(A^T)$, the projection matrix U should have no effect on $\mathbf{y}$. Thus $U\mathbf{y} = \mathbf{y}$. The vector $\mathbf{s} = \mathbf{b} - \mathbf{y}$ is the projection of $\mathbf{b}$ onto $R(A)^{\perp} = N(A)$. Thus $\mathbf{s} \in N(A)$ and $A\mathbf{s} = \mathbf{0}$.

(h) The vectors $\mathbf{s}$ and $V\mathbf{b}$ should be equal since they are both projections of $\mathbf{b}$ onto $N(A)$.

CHAPTER TEST A

1. The statement is false. The statement is true for nonorthogonal vectors, however, if $\mathbf{x} \perp \mathbf{y}$, then the projection of $\mathbf{x}$ onto $\mathbf{y}$ and the projection of $\mathbf{x}$ onto $\mathbf{x}$ are both equal to $\mathbf{0}$.

2. The statement is false. If $\mathbf{x}$ and $\mathbf{y}$ are unit vectors and θ is the angle between the two vectors, then the condition $|\mathbf{x}^T\mathbf{y}| = 1$ implies that $\cos\theta = \pm 1$. Thus $\mathbf{y} = \mathbf{x}$ or $\mathbf{y} = -\mathbf{x}$. So the vectors $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent.

3. The statement is false. For example, consider the one-dimensional subspaces

$$U = \operatorname{Span}(\mathbf{e}_1), \quad V = \operatorname{Span}(\mathbf{e}_3), \quad W = \operatorname{Span}(\mathbf{e}_1 + \mathbf{e}_2)$$

Since $\mathbf{e}_1 \perp \mathbf{e}_3$ and $\mathbf{e}_3 \perp (\mathbf{e}_1 + \mathbf{e}_2)$, it follows that $U \perp V$ and $V \perp W$. However $\mathbf{e}_1$ is not orthogonal to $\mathbf{e}_1 + \mathbf{e}_2$, so U and W are not orthogonal subspaces.

4. The statement is false. If $\mathbf{y}$ is in the column space of and $A^T\mathbf{y} = \mathbf{0}$, then $\mathbf{y}$ is also in $N(A^T)$. But $R(A) \bigcap N(A^T) = \{\mathbf{0}\}$. So $\mathbf{y}$ must be the zero vector.

5. The statement is true. The matrices A and A^TA have the same rank. (See Exercise 13 of Section 2.) Similarly, A^T and AA^T have the same rank. By Theorem 3.6.6 the matrices A and A^T have the same rank. It follows then that

$$\operatorname{rank}(A^TA) = \operatorname{rank}(A) = \operatorname{rank}(A^T) = \operatorname{rank}(AA^T)$$

6. The statement is false. Although the least squares problem will not have a unique solution the projection of a vector onto any subspace is always unique. See Theorem 5.3.1 or Theorem 5.5.8.

7. The statement is true. If A is $m \times n$ and $N(A) = \{\mathbf{0}\}$, then A has rank n and it follows from Theorem 5.3.2 that the least squares problem will have a unique solution.

8. The statement is true. In general an $n \times n$ matrix Q is orthogonal if and only if $Q^TQ = I$. If Q_1 and Q_2 are both $n \times n$ orthogonal matrices, then

$$(Q_1Q_2)^T(Q_1Q_2) = Q_2^TQ_1^TQ_1Q_2 = Q_2^TIQ_2 = Q_2^TQ_2 = I$$

Therefore Q_1Q_2 is an orthogonal matrix.

9. The statement is true. The matrix U^TU is a $k \times k$ and its (i,j) entry is $\mathbf{u}_i^T\mathbf{u}_j$. Since $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are orthonormal vectors, $\mathbf{u}_i^T\mathbf{u}_j = 1$ if $i = j$ and it is equal to 0 otherwise.

10. The statement is false. The statement is only true in the case $k = n$. In the case $k < n$ if we extend the given set of vectors to an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ for R^n and set

$$V = (\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n), \quad W = (U\ V)$$

then W is an orthogonal matrix and

$$I = WW^T = UU^T + VV^T$$

So UU^T is actually equal to $I - VV^T$. As an example let

$$U = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & -\frac{2}{3} \end{pmatrix}$$

The column vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ form an orthonormal set and

$$UU^T = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & -\frac{2}{3} \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{5}{9} & \frac{4}{9} & -\frac{2}{9} \\ \frac{3}{9} & \frac{5}{9} & \frac{2}{9} \\ -\frac{2}{9} & \frac{2}{9} & \frac{8}{9} \end{pmatrix}$$

Thus $UU^T \neq I$. Note that if we set

$$\mathbf{u}_3 = \begin{pmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$$

then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis for R^3 and

$$UU^T + \mathbf{u}_3\mathbf{u}_3^T = \begin{pmatrix} \frac{5}{9} & \frac{4}{9} & -\frac{2}{9} \\ \frac{4}{9} & \frac{5}{9} & \frac{2}{9} \\ -\frac{2}{9} & \frac{2}{9} & \frac{8}{9} \end{pmatrix} + \begin{pmatrix} \frac{4}{9} & -\frac{4}{9} & \frac{2}{9} \\ -\frac{4}{9} & \frac{4}{9} & -\frac{2}{9} \\ \frac{2}{9} & -\frac{2}{9} & \frac{1}{9} \end{pmatrix} = I$$

CHAPTER TEST B

1. **(a)** $\mathbf{p} = \dfrac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y} = \dfrac{3}{9}\mathbf{y} = \left(-\dfrac{2}{3}, \dfrac{1}{3}, \dfrac{2}{3}, 0\right)^T$

(b) $\mathbf{x} - \mathbf{p} = \left(\dfrac{5}{3}, \dfrac{2}{3}, \dfrac{4}{3}, 2\right)^T$

$$(\mathbf{x} - \mathbf{p})^T\mathbf{p} = -\frac{10}{9} + \frac{2}{9} + \frac{8}{9} + 0 = 0$$

(c) $\|\mathbf{x}\|^2 = 1 + 1 + 4 + 4 = 10$

$$\|\mathbf{p}\|^2 + \|\mathbf{x} - \mathbf{p}\|^2 = \left(\frac{4}{9} + \frac{1}{9} + \frac{4}{9} + 0\right) + \left(\frac{25}{9} + \frac{4}{9} + \frac{16}{9} + 4\right) = 1 + 9 = 10$$

2. **(a)** By the Cauchy-Schwarz inequality

$$|\langle \mathbf{v}_1, \mathbf{v}_2 \rangle| \leq \|\mathbf{v}_1\|\|\mathbf{v}_2\|$$

(b) If

$$|\langle \mathbf{v}_1, \mathbf{v}_2 \rangle| = \|\mathbf{v}_1\|\|\mathbf{v}_2\|$$

then equality holds in the Cauchy-Schwarz inequality and this can only happen if the two vectors are linearly dependent.

3.

$$\begin{aligned}
\|\mathbf{v}_1+\mathbf{v}_2\|^2 &= \langle \mathbf{v}_1+\mathbf{v}_2, \mathbf{v}_1+\mathbf{v}_2\rangle \\
&= \langle \mathbf{v}_1, \mathbf{v}_1\rangle + 2\,\langle \mathbf{v}_1, \mathbf{v}_2\rangle + \langle \mathbf{v}_2, \mathbf{v}_2\rangle \\
&\le \|\mathbf{v}_1\|^2 + 2\|\mathbf{v}_1\|\|\mathbf{v}_2\| + \|\mathbf{v}_2\|^2 \quad \text{(Cauchy} - \text{Schwarz)} \\
&= (\|\mathbf{v}_1\| + \|\mathbf{v}_2\|)^2
\end{aligned}$$

4. (a) IfA has rank 4 then A^T must also have rank 4. The matrix A^T has 7 columns, so by the Rank-Nullity theorem its rank and nullity must add up to 7. Since the rank is 4, the nullity must be 3 and hence $\dim N(A^T) = 3$. The orthogonal complement of $N(A^T)$ is $R(A)$.

(b) If $\mathbf{x}$ is in $R(A)$ and $A^T\mathbf{x} = \mathbf{0}$ then $\mathbf{x}$ is also in $N(A^T)$. Since $R(A)$ and $N(A^T)$ are orthogonal subspaces their intersection is $\{\mathbf{0}\}$. Therefore $\mathbf{x} = \mathbf{0}$ and $\|\mathbf{x}\| = 0$.

(c) $\dim N(A^TA) = \dim N(A) = 1$ by the Rank-Nullity Theorem. Therefore the normal equations will involve 1 free variables and hence the least squares problem will have infinitely many solutions.

5. If θ_1 is the angle between $\mathbf{x}$ and $\mathbf{y}$ and θ_2 is the angle between $Q\mathbf{x}$ and $Q\mathbf{y}$ then

$$\cos\theta_2 = \frac{(Q\mathbf{x})^TQ\mathbf{y}}{\|Q\mathbf{x}\|\,\|Q\mathbf{y}\|} = \frac{\mathbf{x}^TQ^TQ\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|} = \cos\theta_1$$

The angles θ_1 and θ_2 must both be in the interval $[0, \pi]$. Since their cosines are equal, the angles must be equal.

6. (a) If we let $X = (\mathbf{x}_1, \mathbf{x}_2)$ then $S = R(X)$ and hence

$$S^\perp = R(X)^\perp = N(X^T)$$

To find a basis for $S^\perp$ we solve $X^T\mathbf{x} = \mathbf{0}$. The matrix

$$X^T = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -2 \end{pmatrix}$$

is already in reduced row echelon form with one free variable $\mathbf{x}_3$. If we set $x_3 = a$, then $x_1 = -2a$ and $x_2 = 2a$. Thus $S^\perp$ consists of all vectors of the form $(-2a, 2a, a)^T$ and $\{(-2, 2, 1)^T\}$ is a basis for $S^\perp$.

(b) S is the span of two linearly independent vectors and hence S can be represented geometrically by a plane through the origin in 3-space. $S^\perp$ corresponds to the line through the original that is normal to the plane representing S.

(c) To find the projection matrix we must find an orthonormal basis for $S^\perp$. Since $\dim S^\perp = 1$ we need only normalize our single basis vector to obtain an orthonormal basis. If we set $\mathbf{u} = \frac{1}{3}(-2, 2, 1)^T$ then the projection matrix is

$$P = \mathbf{u}\mathbf{u}^T = \frac{1}{9}\begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}\begin{pmatrix} -2 & 2 & 1 \end{pmatrix} = \begin{pmatrix} \frac{4}{9} & -\frac{4}{9} & -\frac{2}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{2}{9} \\ -\frac{2}{9} & \frac{2}{9} & \frac{1}{9} \end{pmatrix}$$

7. To find the best least squares fit we must find a least squares solution to the system

$$\begin{aligned} c_1 - c_2 &= 1 \\ c_1 + c_2 &= 3 \\ c_1 + 2c_2 &= 3 \end{aligned}$$

If A is the coefficient matrix for this system and $\mathbf{b}$ is the right hand side, then the solution $\mathbf{c}$ to the least squares problem is the solution to the normal equations $A^TA\mathbf{c} = A^T\mathbf{b}$.

$$A^TA = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}$$

$$A^T\mathbf{b} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 7 \\ 8 \end{pmatrix}$$

The augmented matrix for the normal equations is

$$\left(\begin{array}{cc|c} 3 & 2 & 7 \\ 2 & 6 & 8 \end{array}\right)$$

The solution to this system is $\mathbf{c} = (\frac{13}{7}, \frac{5}{7})^T$ and hence the best linear fit is $f(x) = \frac{13}{7} + \frac{5}{7}x$.

8. (a) It follows from Theorem 5.5.3 that

$$\langle \mathbf{x}, \mathbf{y} \rangle = 2 \cdot 3 + (-2) \cdot 1 + 1 \cdot (-4) = 0$$

(so $\mathbf{x}$ and $\mathbf{y}$ are orthogonal).

(b) By Parseval's formula

$$\|\mathbf{x}\|^2 = 2^2 + (-2)^2 + 1^2 = 9$$

and therefore $\|\mathbf{x}\| = 3$.

9. (a) If $\mathbf{x}$ is any vector in $N(A^T)$ then $\mathbf{x}$ is in $R(A)^\perp$ and hence the projection of $\mathbf{x}$ onto $R(A)$ will be $\mathbf{0}$, i.e., $P\mathbf{x} = \mathbf{0}$. The column vectors of Q are all in $N(A^T)$ since Q projects vectors onto $N(A^T)$ and $\mathbf{q}_j = Q\mathbf{e}_j$ for $1 \le j \le 7$. It follows then that

$$PQ = (P\mathbf{q}_1, P\mathbf{q}_2, P\mathbf{q}_3, P\mathbf{q}_4, P\mathbf{q}_5, P\mathbf{q}_6, P\mathbf{q}_7) = (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}) = O$$

(b) Let $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is an orthonormal basis for $R(A)$ and let $\{\mathbf{u}_5, \mathbf{u}_6, \mathbf{u}_7\}$ be an orthonormal basis for $N(A^T)$. If we set $U_1 = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4)$ and $U_2 = (\mathbf{u}_5, \mathbf{u}_6, \mathbf{u}_7)$ then $P = U_1U_1^T$ and $Q = U_2U_2^T$. The matrix $U = (U_1, U_2)$ is orthogonal and hence $U^{-1} = U^T$. It follows then that

$$I = UU^T = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} U_1^T \\ U_2^T \end{pmatrix} = U_1U_1^T + U_2U_2^T = P + Q$$

10. (a) $r_{13} = \mathbf{q}_1^T\mathbf{a}_3 = -1$, $r_{23} = \mathbf{q}_2^T\mathbf{a}_3 = 3$, $\mathbf{p}_2 = -\mathbf{q}_1 + 3\mathbf{q}_2 = (-2, 1, -2, 1)^T$

$\mathbf{a}_3 - \mathbf{p}_2 = (-3, -3, 3, 3)^T$, $r_{33} = \|\mathbf{a}_3 - \mathbf{p}_2\| = 6$

$\mathbf{q}_3 = \frac{1}{6}(-3, -3, 3, 3)^T = (-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$

(b)

$$\mathbf{c} = Q^T\mathbf{b} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} -6 \\ 1 \\ 1 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \\ 6 \end{pmatrix}$$

To solve the least squares problem we must solve the upper triangular system $R\mathbf{x} = \mathbf{c}$. The augmented matrix for this system is

$$\left(\begin{array}{rrr|r} 2 & -2 & -1 & 1 \\ 0 & 4 & 3 & 6 \\ 0 & 0 & 6 & 6 \end{array}\right)$$

and the solution $\mathbf{x} = (\frac{7}{4}, \frac{3}{4}, 1)^T$ is easily obtained using back substitution.

11. (a) $\langle \cos x, \sin x \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos x \sin x \, dx = 0$

(b) Since $\cos x$ and $\sin x$ are orthogonal we have by the Pythagorean Law that

$$\begin{aligned} \|\cos x + \sin x\|^2 &= \|\cos x\|^2 + \|\sin x\|^2 \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos^2 x \, dx + \frac{1}{\pi}\int_{-\pi}^{\pi} \sin^2 x \, dx \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} 1 \, dx = 2 \end{aligned}$$

Therefore $\|\cos x + \sin x\| = \sqrt{2}$.

12. (a) $\langle u_1(x), u_2(x) \rangle = \int_{-1}^{1} \frac{1}{\sqrt{2}} \frac{\sqrt{6}}{2} x \, dx = 0$

$\langle u_1(x), u_1(x) \rangle = \int_{-1}^{1} \frac{1}{2} \, dx = 1$

$\langle u_2(x), u_2(x) \rangle = \int_{-1}^{1} \frac{3}{2} x^2 \, dx = 1$

(b) Let

$$\begin{aligned} c_1 &= \langle h(x), u_1(x) \rangle = \frac{1}{\sqrt{2}} \int_{-1}^{1} (x^{1/3} + x^{2/3}) \, dx = \frac{6}{5\sqrt{2}} \\ c_2 &= \langle h(x), u_2(x) \rangle = \frac{\sqrt{6}}{2} \int_{-1}^{1} (x^{1/3} + x^{2/3}) x \, dx = \frac{3\sqrt{6}}{7} \end{aligned}$$

The best linear approximation to $h(x)$ is

$$f(x) = c_1 u_1(x) + c_2 u_2(x) = \frac{3}{5} + \frac{9}{7}x$$