AN INTRODUCTION TO OPTIMIZATION

SOLUTIONS MANUAL

Fourth Edition

Edwin K. P. Chong and Stanislaw H. Żak

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

1. Methods of Proof and Some Notation

1.1

A	B	not A	not B	A⇒B	(not B)⇒(not A)
F	F	T	T	T	T
F	T	T	F	T	T
T	F	F	T	F	F
T	T	F	F	T	T

1.2

A	B	not A	not B	A⇒B	not (A and (not B))
F	F	T	T	T	T
F	T	T	F	T	T
T	F	F	T	F	F
T	T	F	F	T	T

1.3

A	B	not (A and B)	not A	not B	(not A) or (not B))
F	F	T	T	T	T
F	T	T	T	F	T
T	F	T	F	T	T
T	T	F	F	F	F

1.4

A	B	A and B	A and (not B)	(A and B) or (A and (not B))
F	F	F	F	F
F	T	F	F	F
T	F	F	T	T
T	T	T	F	T

1.5

The cards that you should turn over are 3 and A. The remaining cards are irrelevant to ascertaining the truth or falsity of the rule. The card with S is irrelevant because S is not a vowel. The card with 8 is not relevant because the rule does not say that if a card has an even number on one side, then it has a vowel on the other side.

Turning over the A card directly verifies the rule, while turning over the 3 card verifies the contraposition.

2. Vector Spaces and Matrices

2.1

We show this by contradiction. Suppose $n < m$. Then, the number of columns of $\boldsymbol{A}$ is n. Since $\operatorname{rank} \boldsymbol{A}$ is the maximum number of linearly independent columns of $\boldsymbol{A}$, then $\operatorname{rank} \boldsymbol{A}$ cannot be greater than $n < m$, which contradicts the assumption that $\operatorname{rank} \boldsymbol{A} = m$.

2.2

⇒: Since there exists a solution, then by Theorem 2.1, $\operatorname{rank} \boldsymbol{A} = \operatorname{rank}[\boldsymbol{A} \vdots \boldsymbol{b}]$. So, it remains to prove that $\operatorname{rank} \boldsymbol{A} = n$. For this, suppose that $\operatorname{rank} \boldsymbol{A} < n$ (note that it is impossible for $\operatorname{rank} \boldsymbol{A} > n$ since $\boldsymbol{A}$ has only n columns). Hence, there exists $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{y} \neq \boldsymbol{0}$, such that $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$ (this is because the columns of

$\boldsymbol{A}$ are linearly dependent, and $\boldsymbol{Ay}$ is a linear combination of the columns of $\boldsymbol{A}$). Let $\boldsymbol{x}$ be a solution to $\boldsymbol{Ax} = \boldsymbol{b}$. Then clearly $\boldsymbol{x} + \boldsymbol{y} \neq \boldsymbol{x}$ is also a solution. This contradicts the uniqueness of the solution. Hence, $\operatorname{rank} \boldsymbol{A} = n$.

$\Leftarrow$: By Theorem 2.1, a solution exists. It remains to prove that it is unique. For this, let $\boldsymbol{x}$ and $\boldsymbol{y}$ be solutions, i.e., $\boldsymbol{Ax} = \boldsymbol{b}$ and $\boldsymbol{Ay} = \boldsymbol{b}$. Subtracting, we get $\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{y}) = \boldsymbol{0}$. Since $\operatorname{rank} \boldsymbol{A} = n$ and $\boldsymbol{A}$ has n columns, then $\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{0}$ and hence $\boldsymbol{x} = \boldsymbol{y}$, which shows that the solution is unique.

2.3

Consider the vectors $\bar{\boldsymbol{a}}_i = [1, \boldsymbol{a}_i^\top]^\top \in \mathbb{R}^{n+1}$, $i = 1, \ldots, k$. Since $k \geq n + 2$, then the vectors $\bar{\boldsymbol{a}}_1, \ldots, \bar{\boldsymbol{a}}_k$ must be linearly independent in $\mathbb{R}^{n+1}$. Hence, there exist $\alpha_1, \ldots \alpha_k$, not all zero, such that

$$\sum_{i=1}^{k} \alpha_i \boldsymbol{a}_i = \boldsymbol{0}.$$

The first component of the above vector equation is $\sum_{i=1}^{k} \alpha_i = 0$, while the last n components have the form $\sum_{i=1}^{k} \alpha_i \boldsymbol{a}_i = \boldsymbol{0}$, completing the proof.

2.4

a. We first postmultiply $\boldsymbol{M}$ by the matrix

$$\begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{O} \\ -\boldsymbol{M}_{m-k,k} & \boldsymbol{I}_{m-k} \end{bmatrix}$$

to obtain

$$\begin{bmatrix} \boldsymbol{M}_{m-k,k} & \boldsymbol{I}_{m-k} \\ \boldsymbol{M}_{k,k} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{O} \\ -\boldsymbol{M}_{m-k,k} & \boldsymbol{I}_{m-k} \end{bmatrix} = \begin{bmatrix} \boldsymbol{O} & \boldsymbol{I}_{m-k} \\ \boldsymbol{M}_{k,k} & \boldsymbol{O} \end{bmatrix}.$$

Note that the determinant of the postmultiplying matrix is 1. Next we postmultiply the resulting product by

$$\begin{bmatrix} \boldsymbol{O} & \boldsymbol{I}_k \\ \boldsymbol{I}_{m-k} & \boldsymbol{O} \end{bmatrix}$$

to obtain

$$\begin{bmatrix} \boldsymbol{O} & \boldsymbol{I}_{m-k} \\ \boldsymbol{M}_{k,k} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{O} & \boldsymbol{I}_k \\ \boldsymbol{I}_{m-k} & \boldsymbol{O} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}_{k,k} \end{bmatrix}.$$

Notice that

$$\det \boldsymbol{M} = \det\left(\begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}_{k,k} \end{bmatrix}\right) \det\left(\begin{bmatrix} \boldsymbol{O} & \boldsymbol{I}_k \\ \boldsymbol{I}_{m-k} & \boldsymbol{O} \end{bmatrix}\right),$$

where

$$\det\left(\begin{bmatrix} \boldsymbol{O} & \boldsymbol{I}_k \\ \boldsymbol{I}_{m-k} & \boldsymbol{O} \end{bmatrix}\right) = \pm 1.$$

The above easily follows from the fact that the determinant changes its sign if we interchange columns, as discussed in Section 2.2. Moreover,

$$\det\left(\begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{M}_{k,k} \end{bmatrix}\right) = \det(\boldsymbol{I}_k) \det(\boldsymbol{M}_{k,k}) = \det(\boldsymbol{M}_{k,k}).$$

Hence,

$$\det \boldsymbol{M} = \pm \det \boldsymbol{M}_{k,k}.$$

b. We can see this on the following examples. We assume, without loss of generality that $\boldsymbol{M}_{m-k,k} = \boldsymbol{O}$ and let $\boldsymbol{M}_{k,k} = 2$. Thus $k = 1$. First consider the case when $m = 2$. Then we have

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{O} & \boldsymbol{I}_{m-k} \\ \boldsymbol{M}_{k,k} & \boldsymbol{O} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}.$$

Thus,

$$\det \boldsymbol{M} = -2 = \det\left(-\boldsymbol{M}_{k,k}\right).$$

Next consider the case when $m = 3$. Then

$$\det \begin{bmatrix} \boldsymbol{O} & \boldsymbol{I}_{m-k} \\ \boldsymbol{M}_{k,k} & \boldsymbol{O} \end{bmatrix} = \det \begin{bmatrix} 0 & \vdots & 1 & 0 \\ 0 & \vdots & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 2 & \vdots & 0 & 0 \end{bmatrix} = 2 \neq \det\left(-\boldsymbol{M}_{k,k}\right).$$

Therefore, in general,

$$\det \boldsymbol{M} \neq \det\left(-\boldsymbol{M}_{k,k}\right)$$

However, when $k = m/2$, that is, when all sub-matrices are square and of the same dimension, then it is true that

$$\det \boldsymbol{M} = \det\left(-\boldsymbol{M}_{k,k}\right).$$

See [121].

2.5

Let

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}$$

and suppose that each block is $k \times k$. John R. Silvester [121] showed that if at least one of the blocks is equal to $\boldsymbol{O}$ (zero matrix), then the desired formula holds. Indeed, if a row or column block is zero, then the determinant is equal to zero as follows from the determinant's properties discussed Section 2.2. That is, if $\boldsymbol{A} = \boldsymbol{B} = \boldsymbol{O}$, or $\boldsymbol{A} = \boldsymbol{C} = \boldsymbol{O}$, and so on, then obviously $\det \boldsymbol{M} = 0$. This includes the case when any three or all four block matrices are zero matrices.

If $\boldsymbol{B} = \boldsymbol{O}$ or $\boldsymbol{C} = \boldsymbol{O}$ then

$$\det \boldsymbol{M} = \det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \det\left(\boldsymbol{A}\boldsymbol{D}\right).$$

The only case left to analyze is when $\boldsymbol{A} = \boldsymbol{O}$ or $\boldsymbol{D} = \boldsymbol{O}$. We will show that in either case,

$$\det \boldsymbol{M} = \det\left(-\boldsymbol{B}\boldsymbol{C}\right).$$

Without loss of generality suppose that $\boldsymbol{D} = \boldsymbol{O}$. Following arguments of John R. Silvester [121], we premultiply $\boldsymbol{M}$ by the product of three matrices whose determinants are unity:

$$\begin{bmatrix} \boldsymbol{I}_k & -\boldsymbol{I}_k \\ \boldsymbol{O} & \boldsymbol{I}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{O} \\ \boldsymbol{I}_k & \boldsymbol{I}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_k & -\boldsymbol{I}_k \\ \boldsymbol{O} & \boldsymbol{I}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{O} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{C} & \boldsymbol{O} \\ \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}.$$

Hence,

$$\begin{aligned} \det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{O} \end{bmatrix} &= \begin{bmatrix} -\boldsymbol{C} & \boldsymbol{O} \\ \boldsymbol{A} & \boldsymbol{B} \end{bmatrix} \\ &= \det\left(-\boldsymbol{C}\right) \det \boldsymbol{B} \\ &= \det\left(-\boldsymbol{I}_k\right) \det \boldsymbol{C} \det \boldsymbol{B}. \end{aligned}$$

Thus we have

$$\det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{O} \end{bmatrix} = \det\left(-\boldsymbol{B}\boldsymbol{C}\right) = \det\left(-\boldsymbol{C}\boldsymbol{B}\right).$$

2.6

We represent the given system of equations in the form $\boldsymbol{Ax} = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 2 & 1 \\ 1 & -2 & 0 & -1 \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Using elementary row operations yields

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 2 & 1 \\ 1 & -2 & 0 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 2 & 1 \\ 0 & -3 & -2 & -2 \end{bmatrix}, \qquad \text{and}$$

$$[\boldsymbol{A}, \boldsymbol{b}] = \begin{bmatrix} 1 & 1 & 2 & 1 & 1 \\ 1 & -2 & 0 & -1 & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 2 & 1 & 1 \\ 0 & -3 & -2 & -2 & -3 \end{bmatrix},$$

from which rank $\boldsymbol{A} = 2$ and rank$[\boldsymbol{A}, \boldsymbol{b}] = 2$. Therefore, by Theorem 2.1, the system has a solution.

We next represent the system of equations as

$$\begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 - 2x_3 - x_4 \\ -2 + x_4 \end{bmatrix}$$

Assigning arbitrary values to x_3 and x_4 ($x_3 = d_3$, $x_4 = d_4$), we get

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 1 - 2x_3 - x_4 \\ -2 + x_4 \end{bmatrix} \\ &= -\frac{1}{3} \begin{bmatrix} -2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 - 2x_3 - x_4 \\ -2 + x_4 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{4}{3}d_3 - \frac{1}{3}d_4 \\ 1 - \frac{2}{3}d_3 - \frac{2}{3}d_4 \end{bmatrix}. \end{aligned}$$

Therefore, a general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -\frac{4}{3}d_3 - \frac{1}{3}d_4 \\ 1 - \frac{2}{3}d_3 - \frac{2}{3}d_4 \\ d_3 \\ d_4 \end{bmatrix} = \begin{bmatrix} -\frac{4}{3} \\ -\frac{2}{3} \\ 1 \\ 0 \end{bmatrix} d_3 + \begin{bmatrix} -\frac{1}{3} \\ -\frac{2}{3} \\ 0 \\ 1 \end{bmatrix} d_4 + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

where d_3 and d_4 are arbitrary values.

2.7

1. Apply the definition of $|-a|$:

$$\begin{aligned} |-a| &= \begin{cases} -a & \text{if } -a > 0 \\ 0 & \text{if } -a = 0 \\ -(-a) & \text{if } -a < 0 \end{cases} \\ &= \begin{cases} -a & \text{if } a < 0 \\ 0 & \text{if } a = 0 \\ a & \text{if } a > 0 \end{cases} \\ &= |a|. \end{aligned}$$

2. If $a \geq 0$, then $|a| = a$. If $a < 0$, then $|a| = -a > 0 > a$. Hence $|a| \geq a$. On the other hand, $|-a| \geq -a$ (by the above). Hence, $a \geq -|-a| = -|a|$ (by property 1).

3. We have four cases to consider. First, if $a, b \geq 0$, then $a + b \geq 0$. Hence, $|a + b| = a + b = |a| + |b|$.
Second, if $a, b \geq 0$, then $a + b \leq 0$. Hence $|a + b| = -(a + b) = -a - b = |a| + |b|$.
Third, if $a \geq 0$ and $b \leq 0$, then we have two further subcases:

1. If $a + b \geq 0$, then $|a + b| = a + b \leq |a| + |b|$.

2. If $a + b \leq 0$, then $|a + b| = -a - b \leq |a| + |b|$.

The fourth case, $a \leq 0$ and $b \geq 0$, is identical to the third case, with a and b interchanged.

4. We first show $|a - b| \leq |a| + |b|$. We have

$$\begin{aligned} |a-b| &= |a+(-b)| \\ &\leq |a| + |-b| \qquad \text{by property 3} \\ &= |a| + |b| \qquad \text{by property 1.} \end{aligned}$$

To show $||a| - |b|| \leq |a - b|$, we note that $|a| = |a - b + b| \leq |a - b| + |b|$, which implies $|a| - |b| \leq |a - b|$. On the other hand, from the above we have $|b| - |a| \leq |b - a| = |a - b|$ by property 1. Therefore, $||a| - |b|| \leq |a - b|$.

5. We have four cases. First, if $a, b \geq 0$, we have $ab \geq 0$ and hence $|ab| = ab = |a||b|$. Second, if $a, b \leq 0$, we have $ab \geq 0$ and hence $|ab| = ab = (-a)(-b) = |a||b|$. Third, if $a \leq 0$, $b \leq 0$, we have $ab \leq 0$ and hence $|ab| = -ab = a(-b) = |a||b|$. The fourth case, $a \leq 0$ and $b \geq 0$, is identical to the third case, with a and b interchanged.

6. We have

$$\begin{aligned} |a+b| &\leq |a| + |b| \qquad \text{by property 3} \\ &\leq c + d. \end{aligned}$$

7. $\Rightarrow$: By property 2, $-a \leq |a|$ and $a \leq |a$. Therefore, $|a| < b$ implies $-a \leq |a| < b$ and $a \leq |a| < b$.
$\Leftarrow$: If $a \geq 0$, then $|a| = a < b$. If $a < 0$, then $|a| = -a < b$.
For the case when "$<$" is replaced by "$\leq$", we simply repeat the above proof with "$<$" replaced by "$\leq$".

8. This is simply the negation of property 7 (apply DeMorgan's Law).

2.8

Observe that we can represent $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_2$ as

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_2 = \boldsymbol{x}^\top \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix} \boldsymbol{y} = (\boldsymbol{Q}\boldsymbol{x})^\top (\boldsymbol{Q}\boldsymbol{y}) = \boldsymbol{x}^\top \boldsymbol{Q}^2 \boldsymbol{y},$$

where

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

Note that the matrix $\boldsymbol{Q} = \boldsymbol{Q}^\top$ is nonsingular.

1. Now, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle_2 = (\boldsymbol{Q}\boldsymbol{x})^\top (\boldsymbol{Q}\boldsymbol{x}) = \|\boldsymbol{Q}\boldsymbol{x}\|^2 \geq 0$, and

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{x} \rangle_2 = 0 \quad &\Leftrightarrow \quad \|\boldsymbol{Q}\boldsymbol{x}\|^2 = 0 \\ &\Leftrightarrow \quad \boldsymbol{Q}\boldsymbol{x} = \boldsymbol{0} \\ &\Leftrightarrow \quad \boldsymbol{x} = \boldsymbol{0} \end{aligned}$$

since $\boldsymbol{Q}$ is nonsingular.

2. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_2 = (\boldsymbol{Q}\boldsymbol{x})^\top (\boldsymbol{Q}\boldsymbol{y}) = (\boldsymbol{Q}\boldsymbol{y})^\top (\boldsymbol{Q}\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle_2$.

3. We have

$$\begin{aligned} \langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle_2 &= (\boldsymbol{x} + \boldsymbol{y})^\top \boldsymbol{Q}^2 \boldsymbol{z} \\ &= \boldsymbol{x}^\top \boldsymbol{Q}^2 \boldsymbol{z} + \boldsymbol{y}^\top \boldsymbol{Q}^2 \boldsymbol{z} \\ &= \langle \boldsymbol{x}, \boldsymbol{z} \rangle_2 + \langle \boldsymbol{y}, \boldsymbol{z} \rangle_2. \end{aligned}$$

4. $\langle r\boldsymbol{x}, \boldsymbol{y}\rangle_2 = (r\boldsymbol{x})^\top \boldsymbol{Q}^2 \boldsymbol{y} = r\boldsymbol{x}^\top \boldsymbol{Q}^2 \boldsymbol{y} = r\langle \boldsymbol{x}, \boldsymbol{y}\rangle_2$.

2.9

We have $\|\boldsymbol{x}\| = \|(\boldsymbol{x}-\boldsymbol{y})+\boldsymbol{y}\| \leq \|\boldsymbol{x}-\boldsymbol{y}\|+\|\boldsymbol{y}\|$ by the Triangle Inequality. Hence, $\|\boldsymbol{x}\|-\|\boldsymbol{y}\| \leq \|\boldsymbol{x}-\boldsymbol{y}\|$. On the other hand, from the above we have $\|\boldsymbol{y}\|-\|\boldsymbol{x}\| \leq \|\boldsymbol{y}-\boldsymbol{x}\| = \|\boldsymbol{x}-\boldsymbol{y}\|$. Combining the two inequalities, we obtain $|\|\boldsymbol{x}\|-\|\boldsymbol{y}\|| \leq \|\boldsymbol{x}-\boldsymbol{y}\|$.

2.10

Let $\epsilon > 0$ be given. Set $\delta = \epsilon$. Hence, if $\|\boldsymbol{x}-\boldsymbol{y}\| < \delta$, then by Exercise 2.9, $|\|\boldsymbol{x}\|-\|\boldsymbol{y}\|| \leq \|\boldsymbol{x}-\boldsymbol{y}\| < \delta = \epsilon$.

3. Transformations

3.1

Let $\boldsymbol{v}$ be the vector such that $\boldsymbol{x}$ are the coordinates of $\boldsymbol{v}$ with respect to $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$, and $\boldsymbol{x}'$ are the coordinates of $\boldsymbol{v}$ with respect to $\{\boldsymbol{e}_1', \boldsymbol{e}_2', \ldots, \boldsymbol{e}_n'\}$. Then,

$$\boldsymbol{v} = x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n]\boldsymbol{x},$$

and

$$\boldsymbol{v} = x_1'\boldsymbol{e}_1' + \cdots + x_n'\boldsymbol{e}_n' = [\boldsymbol{e}_1', \ldots, \boldsymbol{e}_n']\boldsymbol{x}'.$$

Hence,

$$[\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n]\boldsymbol{x} = [\boldsymbol{e}_1', \ldots, \boldsymbol{e}_n']\boldsymbol{x}'$$

which implies

$$\boldsymbol{x}' = [\boldsymbol{e}_1', \ldots, \boldsymbol{e}_n']^{-1}[\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n]\boldsymbol{x} = \boldsymbol{T}\boldsymbol{x}.$$

3.2

a. We have

$$[\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3'] = [\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3]\begin{bmatrix} 1 & 2 & 4 \\ 3 & -1 & 5 \\ -4 & 5 & 3 \end{bmatrix}.$$

Therefore,

$$\boldsymbol{T} = [\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3']^{-1}[\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3] = \begin{bmatrix} 1 & 2 & 4 \\ 3 & -1 & 5 \\ -4 & 5 & 3 \end{bmatrix}^{-1} = \frac{1}{42}\begin{bmatrix} 28 & -14 & -14 \\ 29 & -19 & -7 \\ -11 & 13 & 7 \end{bmatrix}.$$

b. We have

$$[\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3] = [\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3']\begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 0 \\ 3 & 4 & 5 \end{bmatrix}.$$

Therefore,

$$\boldsymbol{T} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 0 \\ 3 & 4 & 5 \end{bmatrix}.$$

3.3

We have

$$[\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3] = [\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3']\begin{bmatrix} 2 & 2 & 3 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \end{bmatrix}.$$

Therefore, the transformation matrix from $\{\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3'\}$ to $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ is

$$\boldsymbol{T} = \begin{bmatrix} 2 & 2 & 3 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \end{bmatrix},$$

Now, consider a linear transformation $L : \mathbb{R}^3 \to \mathbb{R}^3$, and let $\boldsymbol{A}$ be its representation with respect to $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$, and $\boldsymbol{B}$ its representation with respect to $\{\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3'\}$. Let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{y}' = \boldsymbol{B}\boldsymbol{x}'$. Then,

$$\boldsymbol{y}' = \boldsymbol{T}\boldsymbol{y} = \boldsymbol{T}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{T}\boldsymbol{A}(\boldsymbol{T}^{-1}\boldsymbol{x}') = (\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1})\boldsymbol{x}'.$$

Hence, the representation of the linear transformation with respect to $\{\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3'\}$ is

$$\boldsymbol{B} = \boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1} = \begin{bmatrix} 3 & -10 & -8 \\ -1 & 8 & 4 \\ 2 & -13 & -7 \end{bmatrix}.$$

3.4

We have

$$[\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3', \boldsymbol{e}_4'] = [\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4] \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Therefore, the transformation matrix from $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4\}$ to $\{\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3', \boldsymbol{e}_4'\}$ is

$$\boldsymbol{T} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now, consider a linear transformation $L : \mathbb{R}^4 \to \mathbb{R}^4$, and let $\boldsymbol{A}$ be its representation with respect to $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4\}$, and $\boldsymbol{B}$ its representation with respect to $\{\boldsymbol{e}_1', \boldsymbol{e}_2', \boldsymbol{e}_3', \boldsymbol{e}_4'\}$. Let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{y}' = \boldsymbol{B}\boldsymbol{x}'$. Then,

$$\boldsymbol{y}' = \boldsymbol{T}\boldsymbol{y} = \boldsymbol{T}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{T}\boldsymbol{A}(\boldsymbol{T}^{-1}\boldsymbol{x}') = (\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1})\boldsymbol{x}'.$$

Therefore,

$$\boldsymbol{B} = \boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1} = \begin{bmatrix} 5 & 3 & 4 & 3 \\ -3 & -2 & -1 & -2 \\ -1 & 0 & -1 & -2 \\ 1 & 1 & 1 & 4 \end{bmatrix}.$$

3.5

Let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4\}$ be a set of linearly independent eigenvectors of $\boldsymbol{A}$ corresponding to the eigenvalues λ_1, λ_2, λ_3, and λ_4. Let $\boldsymbol{T} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4]$. Then,

$$\begin{aligned} \boldsymbol{A}\boldsymbol{T} &= \boldsymbol{A}[\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4] = [\boldsymbol{A}\boldsymbol{v}_1, \boldsymbol{A}\boldsymbol{v}_2, \boldsymbol{A}\boldsymbol{v}_3, \boldsymbol{A}\boldsymbol{v}_4] \\ &= [\lambda_1\boldsymbol{v}_1, \lambda_2\boldsymbol{v}_2, \lambda_3\boldsymbol{v}_3, \lambda_4\boldsymbol{v}_4] = [\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4] \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix}. \end{aligned}$$

Hence,

$$\boldsymbol{A}\boldsymbol{T} = \boldsymbol{T} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix},$$

or

$$\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}.$$

Therefore, the linear transformation has a diagonal matrix form with respect to the basis formed by a linearly independent set of eigenvectors.

Because

$$\det(\boldsymbol{A}) = (\lambda - 2)(\lambda - 3)(\lambda - 1)(\lambda + 1),$$

the eigenvalues are $\lambda_1 = 2$, $\lambda_2 = 3$, $\lambda_3 = 1$, and $\lambda_4 = -1$.

From $\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$, where $\boldsymbol{v}_i \neq 0$ $(i = 1, 2, 3)$, the corresponding eigenvectors are

$$\boldsymbol{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 2 \\ -9 \\ 1 \end{bmatrix}, \text{and} \qquad \boldsymbol{v}_4 = \begin{bmatrix} 24 \\ -12 \\ 1 \\ 9 \end{bmatrix}.$$

Therefore, the basis we are interested in is

$$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -9 \\ 1 \end{bmatrix}, \begin{bmatrix} 24 \\ -12 \\ 1 \\ 9 \end{bmatrix} \right\}.$$

3.6

Suppose $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are eigenvectors of $\boldsymbol{A}$ corresponding to $\lambda_1, \ldots, \lambda_n$, respectively. Then, for each $i = 1, \ldots, n$, we have

$$(\boldsymbol{I}_n - \boldsymbol{A})\boldsymbol{v}_i = \boldsymbol{v}_i - \boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{v}_i - \lambda_i \boldsymbol{v}_i = (1 - \lambda_i)\boldsymbol{v}_i$$

which shows that $1 - \lambda_1, \ldots, 1 - \lambda_n$ are the eigenvalues of $\boldsymbol{I}_n - \boldsymbol{A}$.

Alternatively, we may write the characteristic polynomial of $\boldsymbol{I}_n - \boldsymbol{A}$ as

$$\pi_{\boldsymbol{I}_n - \boldsymbol{A}}(1 - \lambda) = \det((1 - \lambda)\boldsymbol{I}_n - (\boldsymbol{I}_n - \boldsymbol{A})) = \det(-[\lambda \boldsymbol{I}_n - \boldsymbol{A}]) = (-1)^n \pi_{\boldsymbol{A}}(\lambda),$$

which shows the desired result.

3.7

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}^{\perp}$, and $\alpha, \beta \in \mathbb{R}$. To show that $\mathcal{V}^{\perp}$ is a subspace, we need to show that $\alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{V}^{\perp}$. For this, let $\boldsymbol{v}$ be any vector in $\mathcal{V}$. Then,

$$\boldsymbol{v}^{\top}(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) = \alpha\boldsymbol{v}^{\top}\boldsymbol{x} + \beta\boldsymbol{v}^{\top}\boldsymbol{y} = 0,$$

since $\boldsymbol{v}^{\top}\boldsymbol{x} = \boldsymbol{v}^{\top}\boldsymbol{y} = 0$ by definition.

3.8

The null space of $\boldsymbol{A}$ is $\mathcal{N}(\boldsymbol{A}) = \left\{\boldsymbol{x} \in \mathbb{R}^3 : \boldsymbol{A}\boldsymbol{x} = 0\right\}$. Using elementary row operations and back-substitution, we can solve the system of equations:

$$\begin{bmatrix} 4 & -2 & 0 \\ 2 & 1 & -1 \\ 2 & -3 & 1 \end{bmatrix} \to \begin{bmatrix} 4 & -2 & 0 \\ 0 & 2 & -1 \\ 0 & -2 & 1 \end{bmatrix} \to \begin{bmatrix} 4 & -2 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \Rightarrow \qquad \begin{aligned} 4x_1 - 2x_2 &= 0 \\ 2x_2 - x_3 &= 0 \end{aligned}$$

$$\Rightarrow \qquad x_2 = \frac{1}{2}x_3, \qquad x_1 = \frac{1}{2}x_2 = \frac{1}{4}x_3 \qquad \Rightarrow \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ 1 \end{bmatrix} x_3.$$

Therefore,

$$\mathcal{N}(\boldsymbol{A}) = \left\{ \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} c : c \in \mathbb{R} \right\}.$$

3.9

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$, and $\alpha, \beta \in \mathbb{R}$. Then, there exists $\boldsymbol{v}, \boldsymbol{u}$ such that $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{v}$ and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{u}$. Thus,

$$\alpha\boldsymbol{x} + \beta\boldsymbol{y} = \alpha\boldsymbol{A}\boldsymbol{v} + \beta\boldsymbol{A}\boldsymbol{u} = \boldsymbol{A}(\alpha\boldsymbol{v} + \beta\boldsymbol{u}).$$

Hence, $\alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$, which shows that $\mathcal{R}(\boldsymbol{A})$ is a subspace.

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{N}(\boldsymbol{A})$, and $\alpha, \beta \in \mathbb{R}$. Then, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$. Thus,

$$\boldsymbol{A}(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) = \alpha\boldsymbol{A}\boldsymbol{x} + \beta\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}.$$

Hence, $\alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{N}(\boldsymbol{A})$, which shows that $\mathcal{N}(\boldsymbol{A})$ is a subspace.

3.10

Let $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{B})$, i.e., $\boldsymbol{v} = \boldsymbol{B}\boldsymbol{x}$ for some $\boldsymbol{x}$. Consider the matrix $[\boldsymbol{A}\ \boldsymbol{v}]$. Then, $\mathcal{N}(\boldsymbol{A}^\top) = \mathcal{N}([\boldsymbol{A}\ \boldsymbol{v}]^\top)$, since if $\boldsymbol{u} \in \mathcal{N}(\boldsymbol{A}^\top)$, then $\boldsymbol{u} \in \mathcal{N}(\boldsymbol{B}^\top)$ by assumption, and hence $\boldsymbol{u}^\top\boldsymbol{v} = \boldsymbol{u}^\top\boldsymbol{B}\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{B}^\top\boldsymbol{u} = \boldsymbol{0}$. Now,

$$\dim\mathcal{R}(\boldsymbol{A}) + \dim\mathcal{N}(\boldsymbol{A}^\top) = m$$

and

$$\dim\mathcal{R}([\boldsymbol{A}\ \boldsymbol{v}]) + \dim\mathcal{N}([\boldsymbol{A}\ \boldsymbol{v}]^\top) = m.$$

Since $\dim\mathcal{N}(\boldsymbol{A}^\top) = \dim\mathcal{N}([\boldsymbol{A}\ \boldsymbol{v}]^\top)$, then we have $\dim\mathcal{R}(\boldsymbol{A}) = \dim\mathcal{R}([\boldsymbol{A}\ \boldsymbol{v}])$. Hence, $\boldsymbol{v}$ is a linear combination of the columns of $\boldsymbol{A}$, i.e., $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{A})$, which completes the proof.

3.11

We first show $\boldsymbol{V} \subset (\boldsymbol{V}^\perp)^\perp$. Let $\boldsymbol{v} \in \boldsymbol{V}$, and $\boldsymbol{u}$ any element of $\boldsymbol{V}^\perp$. Then $\boldsymbol{u}^\top\boldsymbol{v} = \boldsymbol{v}^\top\boldsymbol{u} = 0$. Therefore, $\boldsymbol{v} \in (\boldsymbol{V}^\perp)^\perp$.

We now show $(\boldsymbol{V}^\perp)^\perp \subset \boldsymbol{V}$. Let $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ be a basis for $\boldsymbol{V}$, and $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_l\}$ a basis for $(\boldsymbol{V}^\perp)^\perp$. Define $\boldsymbol{A} = [\boldsymbol{a}_1 \cdots \boldsymbol{a}_k]$ and $\boldsymbol{B} = [\boldsymbol{b}_1 \cdots \boldsymbol{b}_l]$, so that $\boldsymbol{V} = \mathcal{R}(\boldsymbol{A})$ and $(\boldsymbol{V}^\perp)^\perp = \mathcal{R}(\boldsymbol{B})$. Hence, it remains to show that $\mathcal{R}(\boldsymbol{B}) \subset \mathcal{R}(\boldsymbol{A})$. Using the result of Exercise 3.10, it suffices to show that $\mathcal{N}(\boldsymbol{A}^\top) \subset \mathcal{N}(\boldsymbol{B}^\top)$. So let $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A}^\top)$, which implies that $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{A})^\perp = \boldsymbol{V}^\perp$, since $\mathcal{R}(\boldsymbol{A})^\perp = \mathcal{N}(\boldsymbol{A}^\top)$. Hence, for all $\boldsymbol{y}$, we have $(\boldsymbol{B}\boldsymbol{y})^\top\boldsymbol{x} = \boldsymbol{0} = \boldsymbol{y}^\top\boldsymbol{B}^\top\boldsymbol{x}$, which implies that $\boldsymbol{B}^\top\boldsymbol{x} = \boldsymbol{0}$. Therefore, $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{B}^\top)$, which completes the proof.

3.12

Let $\boldsymbol{w} \in \mathcal{W}^\perp$, and $\boldsymbol{y}$ be any element of $\mathcal{V}$. Since $\mathcal{V} \subset \mathcal{W}$, then $\boldsymbol{y} \in \mathcal{W}$. Therefore, by definition of $\boldsymbol{w}$, we have $\boldsymbol{w}^\top\boldsymbol{y} = 0$. Therefore, $\boldsymbol{w} \in \mathcal{V}^\perp$.

3.13

Let $r = \dim\mathcal{V}$. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ be a basis for $\mathcal{V}$, and $\boldsymbol{V}$ the matrix whose ith column is $\boldsymbol{v}_i$. Then, clearly $\mathcal{V} = \mathcal{R}(\boldsymbol{V})$.

Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n-r}$ be a basis for $\mathcal{V}^\perp$, and $\boldsymbol{U}$ the matrix whose ith row is $\boldsymbol{u}_i^\top$. Then, $\mathcal{V}^\perp = \mathcal{R}(\boldsymbol{U}^\top)$, and $\mathcal{V} = (\mathcal{V}^\perp)^\perp = \mathcal{R}(\boldsymbol{U}^\top)^\perp = \mathcal{N}(\boldsymbol{U})$ (by Exercise 3.11 and Theorem 3.4).

3.14

---

a. Let $\boldsymbol{x} \in \mathcal{V}$. Then, $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{x} + (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x}$. Note that $\boldsymbol{P}\boldsymbol{x} \in \mathcal{V}$, and $(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x} \in \mathcal{V}^\perp$. Therefore, $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{x} + (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x}$ is an orthogonal decomposition of $\boldsymbol{x}$ with respect to $\mathcal{V}$. However, $\boldsymbol{x} = \boldsymbol{x} + \boldsymbol{0}$ is also an orthogonal decomposition of $\boldsymbol{x}$ with respect to $\mathcal{V}$. Since the orthogonal decomposition is unique, we must have $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{x}$.

b. Suppose $\boldsymbol{P}$ is an orthogonal projector onto $\mathcal{V}$. Clearly, $\mathcal{R}(\boldsymbol{P}) \subset \mathcal{V}$ by definition. However, from part a, $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{V}$, and hence $\mathcal{V} \subset \mathcal{R}(\boldsymbol{P})$. Therefore, $\mathcal{R}(\boldsymbol{P}) = \mathcal{V}$.

3.15

To answer the question, we have to represent the quadratic form with a symmetric matrix as

$$\boldsymbol{x}^\top \left(\frac{1}{2} \begin{bmatrix} 1 & -8 \\ 1 & 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -8 & 1 \end{bmatrix} \right) \boldsymbol{x} = \boldsymbol{x}^\top \begin{bmatrix} 1 & -7/2 \\ -7/2 & 1 \end{bmatrix} \boldsymbol{x}.$$

The leading principal minors are $\Delta_1 = 1$ and $\Delta_2 = -45/4$. Therefore, the quadratic form is indefinite.

3.16

The leading principal minors are $\Delta_1 = 2$, $\Delta_2 = 0$, $\Delta_3 = 0$, which are all nonnegative. However, the eigenvalues of $\boldsymbol{A}$ are $0, -1.4641, 5.4641$ (for example, use Matlab to quickly check this). This implies that the matrix $\boldsymbol{A}$ is indefinite (by Theorem 3.7). An alternative way to show that $\boldsymbol{A}$ is not positive semidefinite is to find a vector $\boldsymbol{x}$ such that $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} < 0$. So, let $\boldsymbol{x}$ be an eigenvector of $\boldsymbol{A}$ corresponding to its negative eigenvalue $\lambda = -1.4641$. Then, $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^\top(\lambda \boldsymbol{x}) = \lambda \boldsymbol{x}^\top \boldsymbol{x} = \lambda \|\boldsymbol{x}\|^2 < 0$. For this example, we can take $\boldsymbol{x} = [0.3251, 0.3251, -0.8881]^\top$, for which we can verify that $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = -1.4643$.

3.17

a. The matrix $\boldsymbol{Q}$ is indefinite, since $\Delta_2 = -1$ and $\Delta_3 = 2$.

b. Let $\boldsymbol{x} \in \mathcal{M}$. Then, $x_2 + x_3 = -x_1$, $x_1 + x_3 = -x_2$, and $x_1 + x_2 = -x_3$. Therefore,

$$\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} = x_1(x_2 + x_3) + x_2(x_1 + x_3) + x_3(x_1 + x_2) = -(x_1^2 + x_2^2 + x_3^2).$$

This implies that the matrix $\boldsymbol{Q}$ is negative definite on the subspace $\mathcal{M}$.

3.18

a. We have

$$f(x_1, x_2, x_3) = x_2^2 = [x_1, x_2, x_3] \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Then,

$$\boldsymbol{Q} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and the eigenvalues of $\boldsymbol{Q}$ are $\lambda_1 = 0$, $\lambda_2 = 1$, and $\lambda_3 = 0$. Therefore, the quadratic form is positive semidefinite.

b. We have

$$f(x_1, x_2, x_3) = x_1^2 + 2x_2^2 - x_1 x_3 = [x_1, x_2, x_3] \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 2 & 0 \\ -\frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Then,

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 2 & 0 \\ -\frac{1}{2} & 0 & 0 \end{bmatrix}$$

and the eigenvalues of $\boldsymbol{Q}$ are $\lambda_1 = 2$, $\lambda_2 = (1 - \sqrt{2})/2$, and $\lambda_3 = (1 + \sqrt{2})/2$. Therefore, the quadratic form is indefinite.

c. We have

$$f(x_1, x_2, x_3) = x_1^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_2x_3 = [x_1, x_2, x_3] \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Then,

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

and the eigenvalues of $\boldsymbol{Q}$ are $\lambda_1 = 0$, $\lambda_2 = 1 - \sqrt{3}$, and $\lambda_3 = 1 + \sqrt{3}$. Therefore, the quadratic form is indefinite.

3.19

We have

$$\begin{aligned} f(x_1,x_2,x_3) &= 4x_1^2+x_2^2+9x_3^2-4x_1x_2-6x_2x_3+12x_1x_3 \\ &= [x_1,x_2,x_3]\begin{bmatrix} 4 & -2 & 6 \\ -2 & 1 & -3 \\ 6 & -3 & 9 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \end{aligned}$$

Let

$$\boldsymbol{Q} = \begin{bmatrix} 4 & -2 & 6 \\ -2 & 1 & -3 \\ 6 & -3 & 9 \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3,$$

where $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, and $\boldsymbol{e}_3$ form the natural basis for $\mathbb{R}^3$.

Let $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$ be another basis for $\mathbb{R}^3$. Then, the vector $\boldsymbol{x}$ is represented in the new basis as $\tilde{\boldsymbol{x}}$, where $\boldsymbol{x} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3]\tilde{\boldsymbol{x}} = \boldsymbol{V}\tilde{\boldsymbol{x}}$.

Now, $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} = (\boldsymbol{V}\tilde{\boldsymbol{x}})^\top \boldsymbol{Q}(\boldsymbol{V}\tilde{\boldsymbol{x}}) = \tilde{\boldsymbol{x}}^\top(\boldsymbol{V}^\top\boldsymbol{Q}\boldsymbol{V})\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}^\top\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{x}}$, where

$$\tilde{\boldsymbol{Q}} = \begin{bmatrix} \tilde{q}_{11} & \tilde{q}_{12} & \tilde{q}_{13} \\ \tilde{q}_{21} & \tilde{q}_{22} & \tilde{q}_{23} \\ \tilde{q}_{31} & \tilde{q}_{32} & \tilde{q}_{33} \end{bmatrix}$$

and $\tilde{q}_{ij} = \boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{v}_j$ for $i,j = 1,2,3$.

We will find a basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ such that $\tilde{q}_{ij} = 0$ for $i \neq j$, and is of the form

$$\begin{aligned} \boldsymbol{v}_1 &= \alpha_{11}\boldsymbol{e}_1 \\ \boldsymbol{v}_2 &= \alpha_{21}\boldsymbol{e}_1 + \alpha_{22}\boldsymbol{e}_2 \\ \boldsymbol{v}_3 &= \alpha_{31}\boldsymbol{e}_1 + \alpha_{32}\boldsymbol{e}_2 + \alpha_{33}\boldsymbol{e}_3 \end{aligned}$$

Because

$$\tilde{q}_{ij} = \boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{v}_j = \boldsymbol{v}_i\boldsymbol{Q}(\alpha_{j1}\boldsymbol{e}_1 + \ldots + \alpha_{jj}\boldsymbol{e}_j) = \alpha_{j1}(\boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{e}_1) + \ldots + \alpha_{jj}(\boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{e}_j),$$

we deduce that if $\boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{e}_j = 0$ for $j < i$, then $\boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{v}_j = 0$. In this case,

$$\tilde{q}_{ii} = \boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{v}_i = \boldsymbol{v}_i\boldsymbol{Q}(\alpha_{i1}\boldsymbol{e}_1 + \ldots + \alpha_{ii}\boldsymbol{e}_i) = \alpha_{i1}(\boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{e}_1) + \ldots + \alpha_{ii}(\boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{e}_i) = \alpha_{ii}(\boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{e}_i).$$

Our task therefore is to find $\boldsymbol{v}_i$ $(i = 1,2,3)$ such that

$$\begin{aligned} \boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{e}_j &= 0, \qquad j < i \\ \boldsymbol{v}_i\boldsymbol{Q}\boldsymbol{e}_i &= 1, \end{aligned}$$

and, in this case, we get

$$\tilde{\boldsymbol{Q}} = \begin{bmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{bmatrix}.$$

- Case of $i = 1$.

  From $\boldsymbol{v}_1^\top\boldsymbol{Q}\boldsymbol{e}_1 = 1$,

  $$(\alpha_{11}\boldsymbol{e}_1)^\top\boldsymbol{Q}\boldsymbol{e}_1 = \alpha_{11}(\boldsymbol{e}_1^\top\boldsymbol{Q}\boldsymbol{e}_1) = \alpha_{11}q_{11} = 1.$$

  Therefore,

  $$\alpha_{11} = \frac{1}{q_{11}} = \frac{1}{\Delta_1} = \frac{1}{4} \qquad \Rightarrow \qquad \boldsymbol{v}_1 = \alpha_{11}\boldsymbol{e}_1 = \begin{bmatrix} \frac{1}{4} \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

- Case of $i = 2$.

  From $\boldsymbol{v}_2^\top \boldsymbol{Q}\boldsymbol{e}_1 = 0$,

$$(\alpha_{21}\boldsymbol{e}_1 + \alpha_{22}\boldsymbol{e}_2)^\top \boldsymbol{Q}\boldsymbol{e}_1 = \alpha_{21}(\boldsymbol{e}_1^\top \boldsymbol{Q}\boldsymbol{e}_1) + \alpha_{22}(\boldsymbol{e}_2^\top \boldsymbol{Q}\boldsymbol{e}_1) = \alpha_{21}q_{11} + \alpha_{22}q_{21} = 0.$$

  From $\boldsymbol{v}_2^\top \boldsymbol{Q}\boldsymbol{e}_2 = 1$,

$$(\alpha_{21}\boldsymbol{e}_1 + \alpha_{22}\boldsymbol{e}_2)^\top \boldsymbol{Q}\boldsymbol{e}_2 = \alpha_{21}(\boldsymbol{e}_1^\top \boldsymbol{Q}\boldsymbol{e}_2) + \alpha_{22}(\boldsymbol{e}_2^\top \boldsymbol{Q}\boldsymbol{e}_2) = \alpha_{21}q_{12} + \alpha_{22}q_{22} = 1.$$

  Therefore,

$$\begin{bmatrix} q_{11} & q_{21} \\ q_{12} & q_{22} \end{bmatrix} \begin{bmatrix} \alpha_{21} \\ \alpha_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

  But, since $\Delta_2 = 0$, this system of equations is inconsistent. Hence, in this problem $\boldsymbol{v}_2^\top \boldsymbol{Q}\boldsymbol{e}_2 = 0$ should be satisfied instead of $\boldsymbol{v}_2^\top \boldsymbol{Q}\boldsymbol{e}_2 = 1$ so that the system can have a solution. In this case, the diagonal matrix becomes

$$\tilde{\boldsymbol{Q}} = \begin{bmatrix} \alpha_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \alpha_{33} \end{bmatrix},$$

  and the system of equations become

$$\begin{bmatrix} q_{11} & q_{21} \\ q_{12} & q_{22} \end{bmatrix} \begin{bmatrix} \alpha_{21} \\ \alpha_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} \alpha_{21} \\ \alpha_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} \alpha_{22},$$

  where α_{22} is an arbitrary real number. Thus,

$$\boldsymbol{v}_2 = \alpha_{21}\boldsymbol{e}_1 + \alpha_{22}\boldsymbol{e}_2 = \begin{bmatrix} \frac{1}{2} \\ 1 \\ 0 \end{bmatrix} a,$$

  where a is an arbitrary real number.

- Case of $i = 3$.

  Since in this case $\Delta_3 = \det(\boldsymbol{Q}) = 0$, we will have to apply the same reasoning of the previous case and use the condition $\boldsymbol{v}_3^\top \boldsymbol{Q}\boldsymbol{e}_3 = 0$ instead of $\boldsymbol{v}_3^\top \boldsymbol{Q}\boldsymbol{e}_3 = 1$. In this way the diagonal matrix becomes

$$\tilde{\boldsymbol{Q}} = \begin{bmatrix} \alpha_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

  Thus, from $\boldsymbol{v}_3^\top \boldsymbol{Q}\boldsymbol{e}_1 = 0$, $\boldsymbol{v}_3^\top \boldsymbol{Q}\boldsymbol{e}_2 = 0$ and $\boldsymbol{v}_3^\top \boldsymbol{Q}\boldsymbol{e}_3 = 0$,

$$\begin{aligned} \begin{bmatrix} q_{11} & q_{21} & q_{31} \\ q_{12} & q_{22} & q_{32} \\ q_{13} & q_{23} & q_{33} \end{bmatrix} \begin{bmatrix} \alpha_{31} \\ \alpha_{32} \\ \alpha_{33} \end{bmatrix} &= \boldsymbol{Q}^\top \begin{bmatrix} \alpha_{31} \\ \alpha_{32} \\ \alpha_{33} \end{bmatrix} = \boldsymbol{Q} \begin{bmatrix} \alpha_{31} \\ \alpha_{32} \\ \alpha_{33} \end{bmatrix} \\ &= \begin{bmatrix} 4 & -2 & 6 \\ -2 & 1 & -3 \\ 6 & -3 & 9 \end{bmatrix} \begin{bmatrix} \alpha_{31} \\ \alpha_{32} \\ \alpha_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \end{aligned}$$

  Therefore,

$$\begin{bmatrix} \alpha_{31} \\ \alpha_{32} \\ \alpha_{33} \end{bmatrix} = \begin{bmatrix} \alpha_{31} \\ 2\alpha_{31} + 3\alpha_{33} \\ \alpha_{33} \end{bmatrix},$$

where α_{31} and α_{33} are arbitrary real numbers. Thus,

$$\boldsymbol{v}_3 = \alpha_{31}\boldsymbol{e}_1 + \alpha_{32}\boldsymbol{e}_2 + \alpha_{33}\boldsymbol{e}_3 = \begin{bmatrix} b \\ 2b+3c \\ c \end{bmatrix},$$

where b and c are arbitrary real numbers.

Finally,

$$\boldsymbol{V} = [x_1, x_2, x_3] = \begin{bmatrix} \frac{1}{4} & \frac{a}{2} & b \\ 0 & a & 2b+3c \\ 0 & 0 & c \end{bmatrix},$$

where a, b, and c are arbitrary real numbers.

3.20

We represent this quadratic form as $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$, where

$$\boldsymbol{Q} = \begin{bmatrix} 1 & \xi & -1 \\ \xi & 1 & 2 \\ -1 & 2 & 5 \end{bmatrix}.$$

The leading principal minors of $\boldsymbol{Q}$ are $\Delta_1 = 1$, $\Delta_2 = 1 - \xi^2$, $\Delta_3 = -5\xi^2 - 4\xi$. For the quadratic form to be positive definite, all the leading principal minors of $\boldsymbol{Q}$ must be positive. This is the case if and only if $\xi \in (-4/5, 0)$.

3.21

The matrix $\boldsymbol{Q} = \boldsymbol{Q}^\top > 0$ can be represented as $\boldsymbol{Q} = \boldsymbol{Q}^{1/2}\boldsymbol{Q}^{1/2}$, where $\boldsymbol{Q}^{1/2} = (\boldsymbol{Q}^{1/2})^\top > 0$.

1. Now, $\langle \boldsymbol{x}, \boldsymbol{x}\rangle_Q = (\boldsymbol{Q}^{1/2}\boldsymbol{x})^\top(\boldsymbol{Q}^{1/2}\boldsymbol{x}) = \|\boldsymbol{Q}^{1/2}\boldsymbol{x}\|^2 \geq 0$, and

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{x}\rangle_Q = 0 \quad &\Leftrightarrow \quad \|\boldsymbol{Q}^{1/2}\boldsymbol{x}\|^2 = 0 \\ &\Leftrightarrow \quad \boldsymbol{Q}^{1/2}\boldsymbol{x} = \boldsymbol{0} \\ &\Leftrightarrow \quad \boldsymbol{x} = \boldsymbol{0} \end{aligned}$$

since $\boldsymbol{Q}^{1/2}$ is nonsingular.

2. $\langle \boldsymbol{x}, \boldsymbol{y}\rangle_Q = \boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{y}^\top\boldsymbol{Q}^\top\boldsymbol{x} = \boldsymbol{y}^\top\boldsymbol{Q}\boldsymbol{x} = \langle \boldsymbol{y}, \boldsymbol{x}\rangle_Q$.
3. We have

$$\begin{aligned} \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z}\rangle_Q &= (\boldsymbol{x}+\boldsymbol{y})^\top\boldsymbol{Q}\boldsymbol{z} \\ &= \boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{z} + \boldsymbol{y}^\top\boldsymbol{Q}\boldsymbol{z} \\ &= \langle \boldsymbol{x}, \boldsymbol{z}\rangle_Q + \langle \boldsymbol{y}, \boldsymbol{z}\rangle_Q. \end{aligned}$$

4. $\langle r\boldsymbol{x}, \boldsymbol{y}\rangle_Q = (r\boldsymbol{x})^\top\boldsymbol{Q}\boldsymbol{y} = r\boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{y} = r\langle \boldsymbol{x}, \boldsymbol{y}\rangle_Q$.

3.22

We have

$$\|\boldsymbol{A}\|_\infty = \max\{\|\boldsymbol{A}\boldsymbol{x}\|_\infty : \|\boldsymbol{x}\|_\infty = 1\}.$$

We first show that $\|\boldsymbol{A}\|_\infty \leq \max_i \sum_{k=1}^n |a_{ik}|$. For this, note that for each $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|_\infty = 1$, we have

$$\begin{aligned} \|\boldsymbol{A}\boldsymbol{x}\|_\infty &= \max_i \left|\sum_{k=1}^n a_{ik}x_k\right| \\ &\leq \max_i \sum_{k=1}^n |a_{ik}||x_k| \\ &\leq \max_i \sum_{k=1}^n |a_{ik}|, \end{aligned}$$

since $|x_k| \leq \max_k |x_k| = \|\boldsymbol{x}\|_\infty = 1$. Therefore,

$$\|\boldsymbol{A}\|_\infty \leq \max_i \sum_{k=1}^{n} |a_{ik}|.$$

To show that $\|\boldsymbol{A}\|_\infty = \max_i \sum_{k=1}^{n} |a_{ik}|$, it remains to find a $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$, $\|\tilde{\boldsymbol{x}}\|_\infty = 1$, such that $\|\boldsymbol{A}\tilde{\boldsymbol{x}}\|_\infty = \max_i \sum_{k=1}^{n} |a_{ik}|$. So, let j be such that

$$\sum_{k=1}^{n} |a_{jk}| = \max_i \sum_{k=1}^{n} |a_{ik}|.$$

Define $\tilde{\boldsymbol{x}}$ by

$$\tilde{x}_k = \begin{cases} |a_{jk}|/a_{jk} & \text{if } a_{jk} \neq 0 \\ 1 & \text{otherwise} \end{cases}.$$

Clearly $\|\tilde{\boldsymbol{x}}\|_\infty = 1$. Furthermore, for $i \neq j$,

$$\left| \sum_{k=1}^{n} a_{ik} \tilde{x}_k \right| \leq \sum_{k=1}^{n} |a_{ik}| \leq \max_i \sum_{k=1}^{n} |a_{ik}| = \sum_{k=1}^{n} |a_{jk}|$$

and

$$\left| \sum_{k=1}^{n} a_{jk} \tilde{x}_k \right| = \sum_{k=1}^{n} |a_{jk}|.$$

Therefore,

$$\|\boldsymbol{A}\tilde{\boldsymbol{x}}\|_\infty = \max_i \left| \sum_{k=1}^{n} a_{ik} \tilde{x}_k \right| = \sum_{k=1}^{n} |a_{jk}| = \max_i \sum_{k=1}^{n} |a_{ik}|.$$

3.23

We have

$$\|\boldsymbol{A}\|_1 = \max\{\|\boldsymbol{A}\boldsymbol{x}\|_1 : \|\boldsymbol{x}\|_1 = 1\}.$$

We first show that $\|\boldsymbol{A}\|_1 \leq \max_k \sum_{i=1}^{m} |a_{ik}|$. For this, note that for each $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|_1 = 1$, we have

$$\begin{aligned} \|\boldsymbol{A}\boldsymbol{x}\|_1 &= \sum_{i=1}^{m} \left| \sum_{k=1}^{n} a_{ik} x_k \right| \\ &\leq \sum_{i=1}^{m} \sum_{k=1}^{n} |a_{ik}||x_k| \\ &\leq \sum_{k=1}^{n} |x_k| \sum_{i=1}^{m} |a_{ik}| \\ &\leq \left(\max_k \sum_{i=1}^{m} |a_{ik}| \right) \sum_{k=1}^{n} |x_k| \\ &\leq \max_k \sum_{i=1}^{m} |a_{ik}|, \end{aligned}$$

since $\sum_{k=1}^{n} |x_k| = \|\boldsymbol{x}\|_1 = 1$. Therefore,

$$\|\boldsymbol{A}\|_1 \leq \max_k \sum_{i=1}^{m} |a_{ik}|.$$

To show that $\|\boldsymbol{A}\|_1 = \max_k \sum_{i=1}^m |a_{ik}|$, it remains to find a $\tilde{\boldsymbol{x}} \in \mathbb{R}^m$, $\|\tilde{\boldsymbol{x}}\|_1 = 1$, such that $\|\boldsymbol{A}\tilde{\boldsymbol{x}}\|_1 = \max_k \sum_{i=1}^m |a_{ik}|$. So, let j be such that

$$\sum_{i=1}^m |a_{ij}| = \max_k \sum_{i=1}^m |a_{ik}|.$$

Define $\tilde{\boldsymbol{x}}$ by

$$\tilde{x}_k = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{otherwise} \end{cases}.$$

Clearly $\|\tilde{\boldsymbol{x}}\|_1 = 1$. Furthermore,

$$\|\boldsymbol{A}\tilde{\boldsymbol{x}}\|_1 = \sum_{i=1}^m \left| \sum_{k=1}^n a_{ik}\tilde{x}_k \right| = \sum_{i=1}^m |a_{ij}| = \max_k \sum_{i=1}^m |a_{ik}|.$$

4. Concepts from Geometry

4.1

$\Rightarrow$: Let $S = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$ be a linear variety. Let $\boldsymbol{x}, \boldsymbol{y} \in S$ and $\alpha \in \mathbb{R}$. Then,

$$\boldsymbol{A}(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) = \alpha\boldsymbol{A}\boldsymbol{x} + (1-\alpha)\boldsymbol{A}\boldsymbol{y} = \alpha\boldsymbol{b} + (1-\alpha)\boldsymbol{b} = \boldsymbol{b}.$$

Therefore, $\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in S$.

$\Leftarrow$: If S is empty, we are done. So, suppose $\boldsymbol{x}_0 \in S$. Consider the set $S_0 = S - \boldsymbol{x}_0 = \{\boldsymbol{x} - \boldsymbol{x}_0 : \boldsymbol{x} \in S\}$. Clearly, for all $\boldsymbol{x}, \boldsymbol{y} \in S_0$ and $\alpha \in \mathbb{R}$, we have $\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in S_0$. Note that $\boldsymbol{0} \in S_0$. We claim that S_0 is a subspace. To see this, let $\boldsymbol{x}, \boldsymbol{y} \in S_0$, and $\alpha \in \mathbb{R}$. Then, $\alpha\boldsymbol{x} = \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{0} \in S_0$. Furthermore, $\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y} \in S_0$, and therefore $\boldsymbol{x} + \boldsymbol{y} \in S_0$ by the previous argument. Hence, S_0 is a subspace. Therefore, by Exercise 3.13, there exists $\boldsymbol{A}$ such that $S_0 = \mathcal{N}(\boldsymbol{A}) = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$. Define $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}_0$. Then,

$$\begin{aligned} S &= S_0 + \boldsymbol{x}_0 = \{\boldsymbol{y} + \boldsymbol{x}_0 : \boldsymbol{y} \in \mathcal{N}(\boldsymbol{A})\} \\ &= \{\boldsymbol{y} + \boldsymbol{x}_0 : \boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}\} \\ &= \{\boldsymbol{y} + \boldsymbol{x}_0 : \boldsymbol{A}(\boldsymbol{y} + \boldsymbol{x}_0) = \boldsymbol{b}\} \\ &= \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}. \end{aligned}$$

4.2

Let $\boldsymbol{u}, \boldsymbol{v} \in \Theta = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| \leq r\}$, and $\alpha \in [0,1]$. Suppose $\boldsymbol{z} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}$. To show that Θ is convex, we need to show that $\boldsymbol{z} \in \Theta$, i.e., $\|\boldsymbol{z}\| \leq r$. To this end,

$$\begin{aligned} \|\boldsymbol{z}\|^2 &= (\alpha\boldsymbol{u}^\top + (1-\alpha)\boldsymbol{v}^\top)(\alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}) \\ &= \alpha^2\|\boldsymbol{u}\|^2 + 2\alpha(1-\alpha)\boldsymbol{u}^\top\boldsymbol{v} + (1-\alpha)^2\|\boldsymbol{v}\|^2. \end{aligned}$$

Since $\boldsymbol{u}, \boldsymbol{v} \in \Theta$, then $\|\boldsymbol{u}\|^2 \leq r^2$ and $\|\boldsymbol{v}\|^2 \leq r^2$. Furthermore, by the Cauchy-Schwarz Inequality, we have $\boldsymbol{u}^\top\boldsymbol{v} \leq \|\boldsymbol{u}\|\|\boldsymbol{v}\| \leq r^2$. Therefore,

$$\|\boldsymbol{z}\|^2 \leq \alpha^2 r^2 + 2\alpha(1-\alpha)r^2 + (1-\alpha)^2 r^2 = r^2.$$

Hence, $\boldsymbol{z} \in \Theta$, which implies that Θ is a convex set, i.e., the any point on the line segment joining $\boldsymbol{u}$ and $\boldsymbol{v}$ is also in Θ.

4.3

Let $\boldsymbol{u}, \boldsymbol{v} \in \Theta = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$, and $\alpha \in [0,1]$. Suppose $\boldsymbol{z} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}$. To show that Θ is convex, we need to show that $\boldsymbol{z} \in \Theta$, i.e., $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}$. To this end,

$$\begin{aligned} \boldsymbol{A}\boldsymbol{z} &= \boldsymbol{A}(\alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}) \\ &= \alpha\boldsymbol{A}\boldsymbol{u} + (1-\alpha)\boldsymbol{A}\boldsymbol{v}. \end{aligned}$$

Since $\boldsymbol{u}, \boldsymbol{v} \in \Theta$, then $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{b}$ and $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{b}$. Therefore,

$$\boldsymbol{A}\boldsymbol{z} = \alpha\boldsymbol{b} + (1-\alpha)\boldsymbol{b} = \boldsymbol{b},$$

and hence $\boldsymbol{z} \in \Theta$.

4.4

Let $\boldsymbol{u}, \boldsymbol{v} \in \Theta = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \geq \boldsymbol{0}\}$, and $\alpha \in [0, 1]$. Suppose $\boldsymbol{z} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}$. To show that Θ is convex, we need to show that $\boldsymbol{z} \in \Theta$, i.e., $\boldsymbol{z} \geq \boldsymbol{0}$. To this end, write $\boldsymbol{x} = [x_1, \ldots, x_n]^\top$, $\boldsymbol{y} = [y_1, \ldots, y_n]^\top$, and $\boldsymbol{z} = [z_1, \ldots, z_n]^\top$. Then, $z_i = \alpha x_i + (1-\alpha)y_i$, $i = 1, \ldots, n$. Since $x_i, y_i \geq 0$, and $\alpha, 1-\alpha \geq 0$, we have $z_i \geq 0$. Therefore, $\boldsymbol{z} \geq \boldsymbol{0}$, and hence $\boldsymbol{z} \in \Theta$.

5. Elements of Calculus

5.1

Observe that

$$\|\boldsymbol{A}^k\| \leq \|\boldsymbol{A}^{k-1}\|\|\boldsymbol{A}\| \leq \|\boldsymbol{A}^{k-2}\|\|\boldsymbol{A}\|^2 \leq \cdots \leq \|\boldsymbol{A}\|^k.$$

Therefore, if $\|\boldsymbol{A}\| < 1$, then $\lim_{k\to\infty} \|\boldsymbol{A}^k\| = \boldsymbol{O}$ which implies that $\lim_{k\to\infty} \boldsymbol{A}^k = \boldsymbol{O}$.

5.2

For the case when $\boldsymbol{A}$ has all real eigenvalues, the proof is simple. Let λ be the eigenvalue of $\boldsymbol{A}$ with largest absolute value, and $\boldsymbol{x}$ the corresponding (normalized) eigenvector, i.e., $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ and $\|\boldsymbol{x}\| = 1$. Then,

$$\|\boldsymbol{A}\| \geq \|\boldsymbol{A}\boldsymbol{x}\| = \|\lambda\boldsymbol{x}\| = |\lambda|\|\boldsymbol{x}\| = |\lambda|,$$

which completes the proof for this case.

In general, the eigenvalues of $\boldsymbol{A}$ and the corresponding eigenvectors may be complex. In this case, we proceed as follows (see [41]). Consider the matrix

$$\boldsymbol{B} = \frac{\boldsymbol{A}}{\|\boldsymbol{A}\| + \varepsilon},$$

where ε is a positive real number. We have

$$\|\boldsymbol{B}\| = \frac{\|\boldsymbol{A}\|}{\|\boldsymbol{A}\| + \varepsilon} < 1.$$

By Exercise 5.1, $\boldsymbol{B}^k \to \boldsymbol{O}$ as $k \to \infty$, and thus by Lemma 5.1, $|\lambda_i(\boldsymbol{B})| < 1$, $i = 1, \ldots, n$. On the other hand, for each $i = 1, \ldots, n$,

$$\lambda_i(\boldsymbol{B}) = \frac{\lambda_i(\boldsymbol{A})}{\|\boldsymbol{A}\| + \varepsilon},$$

and thus

$$|\lambda_i(\boldsymbol{B})| = \frac{|\lambda_i(\boldsymbol{A})|}{\|\boldsymbol{A}\| + \varepsilon} < 1.$$

which gives

$$|\lambda_i(\boldsymbol{A})| < \|\boldsymbol{A}\| + \varepsilon.$$

Since the above arguments hold for any $\varepsilon > 0$, we have $|\lambda_i(\boldsymbol{A})| \leq \|\boldsymbol{A}\|$.

5.3

a. $\nabla f(\boldsymbol{x}) = (\boldsymbol{a}\boldsymbol{b}^\top + \boldsymbol{b}\boldsymbol{a}^\top)\boldsymbol{x}$.

b. $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{a}\boldsymbol{b}^\top + \boldsymbol{b}\boldsymbol{a}^\top$.

5.4

We have

$$Df(\boldsymbol{x}) = [x_1/3, x_2/2],$$

and

$$\frac{d\boldsymbol{g}}{dt}(t) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

By the chain rule,

$$\begin{aligned} \frac{d}{dt}F(t) &= Df(\boldsymbol{g}(t))\frac{d\boldsymbol{g}}{dt}(t) \\ &= [(3t+5)/3, (2t-6)/2] \begin{bmatrix} 3 \\ 2 \end{bmatrix} \\ &= 5t - 1. \end{aligned}$$

5.5

We have

$$Df(\boldsymbol{x}) = [x_2/2, x_1/2],$$

and

$$\frac{\partial \boldsymbol{g}}{\partial s}(s,t) = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \qquad \frac{\partial \boldsymbol{g}}{\partial t}(s,t) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

By the chain rule,

$$\begin{aligned} \frac{\partial}{\partial s} f(\boldsymbol{g}(s,t)) &= Df(\boldsymbol{g}(t))\frac{\partial \boldsymbol{g}}{\partial s}(s,t) \\ &= \frac{1}{2}[2s+t, 4s+3t] \begin{bmatrix} 4 \\ 2 \end{bmatrix} \\ &= 8s + 5t, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial t} f(\boldsymbol{g}(s,t)) &= Df(\boldsymbol{g}(t))\frac{\partial \boldsymbol{g}}{\partial t}(s,t) \\ &= \frac{1}{2}[2s+t, 4s+3t] \begin{bmatrix} 3 \\ 1 \end{bmatrix} \\ &= 5s + 3t. \end{aligned}$$

5.6

We have

$$Df(\boldsymbol{x}) = [3x_1^2x_2x_3^2 + x_2,\ x_1^3x_3^2 + x_1,\ 2x_1^3x_2x_3 + 1]$$

and

$$\frac{d\boldsymbol{x}}{dt}(t) = \begin{bmatrix} e^t + 3t^2 \\ 2t \\ 1 \end{bmatrix}.$$