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 rank $\boldsymbol{A}$ is the maximum number of linearly independent columns of $\boldsymbol{A}$, then rank $\boldsymbol{A}$ cannot be greater than $n < m$, which contradicts the assumption that rank $\boldsymbol{A} = m$.

2.2

$\Rightarrow$: Since there exists a solution, then by Theorem 2.1, rank $\boldsymbol{A} = \text{rank}[\boldsymbol{A} \vdots \boldsymbol{b}]$. So, it remains to prove that rank $\boldsymbol{A} = n$. For this, suppose that rank $\boldsymbol{A} < n$ (note that it is impossible for 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 $\boldsymbol{A}\boldsymbol{y}$ is a linear combination of the columns of $\boldsymbol{A}$). Let $\boldsymbol{x}$ be a solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. Then clearly $\boldsymbol{x} + \boldsymbol{y} \neq \boldsymbol{x}$ is also a solution. This contradicts the uniqueness of the solution. Hence, 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., $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ and $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}$. Subtracting, we get $\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{y}) = \boldsymbol{0}$. Since 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 $\bar{\boldsymbol{a}}_i = [1, \boldsymbol{a}_i^T]^T \in \mathbb{R}^{n+1}$, $i = 1, \ldots, k$. 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

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

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

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

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

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).

2.5

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

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

where

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

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

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

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

since $\boldsymbol{Q}$ is nonsingular.

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

3. We have

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

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

2.6

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 $|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|| \leq \|\boldsymbol{x} - \boldsymbol{y}\|$.

2.7

Let $\epsilon > 0$ be given. Set $\delta = \epsilon$. Hence, if $\|\boldsymbol{x} - \boldsymbol{y}\| < \delta$, then by Exercise 2.6, $|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|| \leq \|\boldsymbol{x} - \boldsymbol{y}\| < \delta = \epsilon$.

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,

$$[e_1, \ldots, e_n]x = [e'_1, \ldots, e'_n]x'$$

which implies

$$x' = [e'_1, \ldots, e'_n]^{-1}[e_1, \ldots, e_n]x = Tx.$$

3.2

Suppose $v_1, \ldots, v_n$ are eigenvectors of A corresponding to $\lambda_1, \ldots, \lambda_n$, respectively. Then, for each $i = 1, \ldots, n$, we have

$$(I_n - A)v_i = v_i - Av_i = v_i - \lambda v_i = (1 - \lambda_i)v_i$$

which shows that $1 - \lambda_1, \ldots, 1 - \lambda_n$ are the eigenvalues of $I_n - A$.

Alternatively, we may write the characteristic polynomial of $I_n - A$ as

$$\pi_{I_n - A}(1 - \lambda) = \det((1 - \lambda)I_n - (I_n - A)) = \det(-[\lambda I_n - A]) = (-1)^n \pi_A(\lambda),$$

which shows the desired result.

3.3

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

$$v^T(\alpha x + \beta y) = \alpha v^T x + \beta v^T y = 0,$$

since $v^T x = v^T y = 0$ by definition.

3.4

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

$$\alpha x + \beta y = \alpha Av + \beta Au = A(\alpha v + \beta u).$$

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

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

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay = 0.$$

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

3.5

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

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

and

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

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

3.6

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

We now show $(V^\perp)^\perp \subset V$. Let $\{a_1, \ldots, a_k\}$ be a basis for V, and $\{b_1, \ldots, b_l\}$ a basis for $(V^\perp)^\perp$. Define $A = [a_1 \cdots a_k]$ and $B = [b_1 \cdots b_l]$, so that $V = \mathcal{R}(A)$ and $(V^\perp)^\perp = \mathcal{R}(B)$. Hence, it remains to show that $\mathcal{R}(B) \subset \mathcal{R}(A)$. Using the result of Exercise 3.5, it suffices to show that $\mathcal{N}(A^T) \subset \mathcal{N}(B^T)$. So let $x \in \mathcal{N}(A^T)$, which implies that $x \in \mathcal{R}(A)^\perp = V^\perp$, since $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$. Hence, for all y, we have $(By)^T x = 0 = y^T B^T x$, which implies that $B^T x = 0$. Therefore, $x \in \mathcal{N}(B^T)$, which completes the proof.

3.7

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

3.8

Let $r = \dim \mathcal{V}$. Let $v_1, \ldots, v_r$ be a basis for $\mathcal{V}$, and V the matrix whose ith column is v_i. Then, clearly $\mathcal{V} = \mathcal{R}(V)$.

Let $u_1, \ldots, u_{n-r}$ be a basis for $\mathcal{V}^\perp$, and U the matrix whose ith row is u_i^T. Then, $\mathcal{V}^\perp = \mathcal{R}(U^T)$, and $\mathcal{V} = (\mathcal{V}^\perp)^\perp = \mathcal{R}(U^T)^\perp = \mathcal{N}(U)$ (by Exercise 3.6 and Theorem 3.4).

3.9

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

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

3.10

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

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

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

3.11

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

3.12

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

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

$$x^T Q x = x_1(x_2 + x_3) + x_2(x_1 + x_3) + x_3(x_1 + x_2) = -(x_1^2 + x_2^2 + x_3^2).$$

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

3.13

We represent this quadratic form as $f(x) = x^T Q x$, where

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

The leading principal minors of 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 Q must be positive. This is the case if and only if $\xi \in (-4/5, 0)$.

3.14

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

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

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

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

2. $\langle x, y \rangle_Q = x^T Q y = y^T Q^T x = y^T Q x = \langle y, x \rangle_Q$.

3. We have

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

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

3.15

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,

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

3.16

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

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

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

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

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

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

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

Define $\tilde{x}$ by

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

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

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

4. Concepts from Geometry

4.1

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

$$
A(\alpha x + (1-\alpha)y) = \alpha Ax + (1-\alpha)Ay = \alpha b + (1-\alpha)b = b.
$$

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

$\Leftarrow$: If S is empty, we are done. So, suppose $x_0 \in S$. Consider the set $S_0 = S - x_0 = \{x - x_0 : x \in S\}$. Clearly, for all $x, y \in S_0$ and $\alpha \in \mathbb{R}$, we have $\alpha x + (1-\alpha)y \in S_0$. Note that $0 \in S_0$. We claim that S_0 is a subspace. To see this, let $x, y \in S_0$, and $\alpha \in \mathbb{R}$. Then, $\alpha x = \alpha x + (1-\alpha)0 \in S_0$. Furthermore, $\frac{1}{2}x + \frac{1}{2}y \in S_0$, and therefore $x + y \in S_0$ by the previous argument. Hence, S_0 is a subspace. Therefore, by Exercise 3.8, there exists A such that $S_0 = \mathcal{N}(A) = \{x : Ax = 0\}$. Define $b = Ax_0$. Then,

$$
\begin{aligned}
S &= S_0 + x_0 = \{y + x_0 : y \in \mathcal{N}(A)\} \\
&= \{y + x_0 : Ay = 0\} \\
&= \{y + x_0 : A(y + x_0) = b\} \\
&= \{x : Ax = b\}.
\end{aligned}
$$

4.2

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

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

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

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

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

4.3

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

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

Since $u, v \in \Theta$, then $Au = b$ and $Av = b$. Therefore,

$$Az = \alpha b + (1-\alpha) b = b,$$

and hence $z \in \Theta$.

4.4

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

5. Elements of Calculus

5.1

Observe that

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

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

5.2

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

$$\|A\| \geq \|Ax\| = \|\lambda x\| = |\lambda|\|x\| = |\lambda|,$$

which completes the proof for this case.

In general, the eigenvalues of A and the corresponding eigenvectors may be complex. In this case, we proceed as follows (see [27]). Consider the matrix

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

where ε is a positive real number. We have

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

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

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

and thus

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

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

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,

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

5.4

We have

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

and

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

By the chain rule,

$$\begin{aligned} \frac{\partial}{\partial s}f(\boldsymbol{g}(s,t)) &= Df(\boldsymbol{g}(t))\frac{\partial \boldsymbol{g}}{\partial s}(s,t) \\ &= \frac{1}{2}[2s+t, 4s+3t]\begin{bmatrix} 2 \\ 4 \end{bmatrix} \\ &= 10s + 7t, \end{aligned}$$

and

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

5.5

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

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

By the chain rule,

$$\begin{aligned} \frac{d}{dt}f(\boldsymbol{x}(t)) &= Df(\boldsymbol{x}(t))\frac{d\boldsymbol{x}}{dt}(t) \\ &= [3x_1(t)^2x_2(t)x_3(t)^2 + x_2(t),\ x_1(t)^3x_3(t)^2 + x_1(t),\ 2x_1(t)^3x_2(t)x_3(t) + 1]\begin{bmatrix} e^t + 3t^2 \\ 2t \\ 1 \end{bmatrix} \\ &= 12t(e^t + 3t^2)^3 + 2te^t + 6t^2 + 2t + 1. \end{aligned}$$

5.6

Let $\varepsilon > 0$ be given. Since $\boldsymbol{f}(\boldsymbol{x}) = o(g(\boldsymbol{x}))$, then

$$\lim_{\boldsymbol{x}\to\mathbf{0}} \frac{\|\boldsymbol{f}(\boldsymbol{x})\|}{g(\boldsymbol{x})} = 0.$$

Hence, there exists $\delta > 0$ such that if $\|\boldsymbol{x}\| < \delta$, then

$$\frac{\|\boldsymbol{f}(\boldsymbol{x})\|}{g(\boldsymbol{x})} < \varepsilon,$$

which can be rewritten as

$$\|\boldsymbol{f}(\boldsymbol{x})\| \le \varepsilon g(\boldsymbol{x}).$$

5.7

By Exercise 5.6, there exists $\delta > 0$ such that if $\|\boldsymbol{x}\| < \delta$, then $|o(g(\boldsymbol{x}))| < g(\boldsymbol{x})/2$. Hence, if $\|\boldsymbol{x}\| < \delta$, $\boldsymbol{x} \neq \mathbf{0}$, then

$$f(\boldsymbol{x}) \le -g(\boldsymbol{x}) + |o(g(\boldsymbol{x}))| < -g(\boldsymbol{x}) + g(\boldsymbol{x})/2 = -\frac{1}{2}g(\boldsymbol{x}) < 0.$$

5.8

We have that

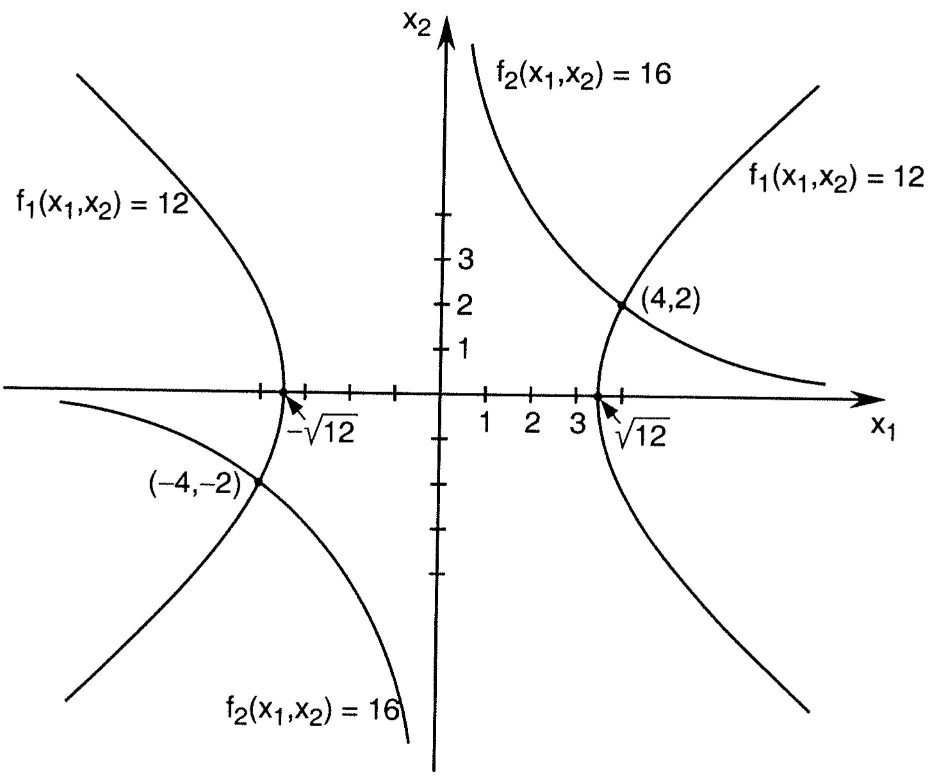
$$\{\boldsymbol{x} : f_1(\boldsymbol{x}) = 12\} = \{\boldsymbol{x} : x_1^2 - x_2^2 = 12\},$$

and

$$\{\boldsymbol{x} : f_2(\boldsymbol{x}) = 16\} = \{\boldsymbol{x} : x_2 = 8/x_1\}.$$

To find the intersection points, we substitute $x_2 = 8/x_1$ into $x_1^2 - x_2^2 = 12$ to get $x_1^4 - 12x_1^2 - 64 = 0$. Solving gives $x_1^2 = 16, -4$. Clearly, the only two possibilities for x_1 are $x_1 = +4, -4$, from which we obtain $x_2 = +2, -2$. Hence, the intersection points are located at $[4, 2]^T$ and $[-4, -2]^T$.

The level sets associated with $f_1(x_1, x_2) = 12$ and $f_2(x_1, x_2) = 16$ are shown as follows.



5.9

a. We have

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_o) + Df(\boldsymbol{x}_o)(\boldsymbol{x} - \boldsymbol{x}_o) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_o)^T D^2 f(\boldsymbol{x}_o)(\boldsymbol{x} - \boldsymbol{x}_o) + \cdots.$$

We compute

$$\begin{aligned} Df(\boldsymbol{x}) &= [e^{-x_2}, -x_1 e^{-x_2} + 1], \\ D^2 f(\boldsymbol{x}) &= \begin{bmatrix} 0 & -e^{-x_2} \\ -e^{-x_2} & x_1 e^{-x_2} \end{bmatrix}. \end{aligned}$$

Hence,

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

b. We compute

$$\begin{aligned} Df(\boldsymbol{x}) &= [4x_1^3 + 4x_1 x_2^2, 4x_1^2 x_2 + 4x_2^3], \\ D^2 f(\boldsymbol{x}) &= \begin{bmatrix} 12x_1^2 + 4x_2^2 & 8x_1 x_2 \\ 8x_1 x_2 & 4x_1^2 + 12x_2^2 \end{bmatrix}. \end{aligned}$$

Expanding f about the point $\boldsymbol{x}_o$ yields

$$\begin{aligned} f(\boldsymbol{x}) &= 4 + [8, 8] \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} + \frac{1}{2}[x_1 - 1, x_2 - 1] \begin{bmatrix} 16 & 8 \\ 8 & 16 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} + \cdots \\ &= 8x_1^2 + 8x_2^2 - 16x_1 - 16x_2 + 8x_1 x_2 + 12 + \cdots. \end{aligned}$$

c. We compute

$$\begin{aligned} Df(\boldsymbol{x}) &= [e^{x_1 - x_2} + e^{x_1 + x_2} + 1, -e^{x_1 - x_2} + e^{x_1 + x_2} + 1], \\ D^2 f(\boldsymbol{x}) &= \begin{bmatrix} e^{x_1 - x_2} + e^{x_1 + x_2} & -e^{x_1 - x_2} + e^{x_1 + x_2} \\ -e^{x_1 - x_2} + e^{x_1 + x_2} & e^{x_1 - x_2} + e^{x_1 + x_2} \end{bmatrix}. \end{aligned}$$

Expanding f about the point $\boldsymbol{x}_o$ yields

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