Contents

Chapter 1

Linear algebra

1.1 A recollection of basic concepts

Problem 1.1 (solution 1.1)

Determine the eigenvalues and corresponding eigenvectors of the matrix

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

Problem 1.2 (solution 1.2)

Determine the eigenvalues and orthogonal bases corresponding eigenvectors of the matrix

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 1 \end{bmatrix}$$

Problem 1.3 (solution 1.3)

Let A be an $n \times n$ matrix. One way to define the exponential of the matrix A is

$$e^A := \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$

(we substituted $x \in \mathbb{R}$ with A in the Taylor series for the function e^x).

(a) Prove the identity

$$\det(e^A) = e^{\operatorname{tr}(A)}$$

(b) Assume A is an *asymmetrical* matrix, meaning $A^\top = -A$. Show that e^A is an orthogonal matrix with determinant equal to 1.

Problem 1.4 (solution 1.4)

You are given an $n \times n$ matrix

$$A = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix},$$

ie. the adjacency matrix of a (undirected) star graph.

(a) Determine the bases of $N(A)$ and $C(A)$, i.e. bases of the null space and the column space of A.

(b) Determine the eigenvalues and eigenvectors of A.
Hint: Why is $N(A)$ an eigenspace of A? Why is $N(A)^{\perp}$ the sum of other eigenspaces of A?

Problem 1.5 (solution 1.5)

The following is known about a *symmetric* matrix $A \in \mathbb{R}^{4\times 4}$: A has 3 as a double eigenvalue and it interchanges the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

i.e. $A\mathbf{v}_1 = \mathbf{v}_2$ and $A\mathbf{v}_2 = \mathbf{v}_1$. Find such a matrix A or prove that it does not exist!

1.2 Schur decomposition, Frobenius norm, Eckart–Young theorem

Problem 1.6 (solution 1.6)

Determine the Schur decompositions of matrices

$$A = \begin{bmatrix} 6 & -1 & 1 \\ 4 & 3 & 1 \\ 2 & 2 & 3 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 0 \\ -\sqrt{2} & -\sqrt{2} & 2 \end{bmatrix}.$$

Problem 1.7 (solution 1.7)

Let A be an arbitrary matrix and let U and V be orthogonal matrices, so that one can form the product UAV. Prove that the following equalities hold:

1. $\|UA\|_{\mathrm{F}} = \|A\|_{\mathrm{F}}$,
2. $\|AV\|_{\mathrm{F}} = \|A\|_{\mathrm{F}}$,

3. $\|UAV\|_F = \|A\|_F$.

Problem 1.8 (solution 1.8)

Denote by $\langle A, B\rangle_F := \operatorname{tr}(A^\top B)$ the Frobenius inner product of matrices $A, B \in \mathbb{R}^{m\times n}$, and denote by $\|A\|_F := \sqrt{\langle A, A\rangle_F}$ the corresponding Frobenius norm. Prove:

1. *the Cauchy–Schwarz inequality,* $|\langle A, B\rangle_F| \le \|A\|_F \|B\|_F$,
2. *the triangle inequality,* $\|A + B\|_F \le \|A\|_F + \|B\|_F$.
3. *submultiplicativity,* $\|AB\|_F \le \|A\|_F \|B\|_F$.
4. *multiplicativity for the Kronecker product,* $\|A \otimes B\|_F = \|A\|_F \|B\|_F$

Problem 1.9 (solution 1.9)

Let I be the 2×2 identity matrix. Find an orthonormal basis of orthogonal complement of I, $I^\perp \le \mathbb{R}^{2\times 2}$, with respect to the (Frobenius) inner product $\langle A, B\rangle_F := \operatorname{tr}(A^\top B)$.

Problem 1.10 (solution 1.10)

Find rank 1 matrices closest (with respect to the Frobenius norm) to the matrices:

(a) $\begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, (b) $\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$, (c) $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

Are those rank 1 matrices unique?

Problem 1.11 (solution 1.11)

Let $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{n\times n}$. Show that the *Kronecker sum*

$$A \oplus B := A \otimes I_n + I_m \otimes B$$

has the property: Eigenvalues of $A \oplus B$ are all possible sums of the form $\lambda_i + \mu_j$, where $\lambda_1, \dots, \lambda_m$ are eigenvalues of A, and $\mu_1, \dots, \mu_n$ eigenvalues of B.

Use this to find eigenvalues and eigenvectors of $A \oplus B$, where

$$A = \begin{bmatrix} -1 & 2 \\ 0 & 3 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}.$$

Problem 1.12 (solution 1.12)

Let

$$A = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}.$$

(a) Determine a diagonal matrix D and an orthogonal matrix U such that $A = UDU^\mathsf{T}$.

(b) Explain why for an orthogonal matrix U the matrix $U \otimes U$ is also orthogonal.

(c) Find matrices of rank 1 and 2 which are best approximations to the Kronecker product $A \otimes A$ w.r.t. the Frobenius norm.

Problem 1.13 (solution 1.13)

Find the eigenvalues and corresponding eigenvectors of the matrix $A \otimes A + A^2 \otimes I$, where A is the matrix

$$A = \begin{bmatrix} -1 & 3 \\ 3 & -1 \end{bmatrix}.$$

Problem 1.14 (solution 1.14)

The objective of this exercise is to express the Sylvester matrix equation $AX + XB = C$ in the 'usual' form ($\hat{A}\mathbf{x} = \mathbf{b}$) using the vec operator and then solve this equation.

(a) Confirm that the matrix equation $AX + XB = C$ in the unknown matrix X is equivalent to the linear system

$$(B^\mathsf{T} \oplus A)\text{vec}(X) = \text{vec}(C)$$

in the unknown column $\text{vec}(X)$.

(b) Let A and B be 2×2 matrices

$$A = \begin{bmatrix} -1 & 2 \\ 0 & 3 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}.$$

Does $AX + XB = 0$ posses a non-trivial solution? (You need to answer quickly! Do not attempt to solve the corresponding linear system...)

(c) Find a matrix X which solves

$$AX + XB = \begin{bmatrix} -2 & 1 \\ 2 & 5 \end{bmatrix}.$$

Problem 1.15 (solution 1.15)

Let $A \in \mathbb{R}^{n \times n}$ be a matrix with only nonnegative eigenvalues.

(a) Prove that A is invertible if and only if all of its eigenvalues are (strictly) positive.

(b) Assume A is invertible. Prove that all of its eigenvalues of A are positive if and only if all of the eigenvalues of A^{-1} are positive.

(c) Assume $A^\mathsf{T} = A$. Prove that there exists a matrix S, with only nonnegative eigenvalues and $S^2 = A$ holds. We denote such matrix S as $S = \sqrt{A}$.

Problem 1.16 (solution 1.16)

You are given the matrix

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 6 & 3 \\ 1 & 3 & 2 \end{bmatrix}.$$

(a) Check that A is positive semidefinite.

(b) Find all eigenvalues and corresponding eigenvectors of A.

(c) Find $\sqrt{A}$.

Problem 1.17 (solution 1.17)

Find the Cholesky decomposition ($A = LL^\mathsf{T}$, where L is lower triangular) of the matrix

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 8 & 2 \\ -1 & 2 & 6 \end{bmatrix}$$

using (recursive) algorithm below:
Write a symmetric matrix $A \in \mathbb{R}^{n\times n}$ in the block form

$$A_1 := A = \begin{bmatrix} a_{11} & \mathbf{b}^\mathsf{T} \\ \mathbf{b} & B \end{bmatrix}$$

and define

$$L_1 := \begin{bmatrix} \sqrt{a_{11}} & \mathbf{0}^\mathsf{T} \\ \frac{1}{\sqrt{a_{11}}}\mathbf{b} & I_{n-1} \end{bmatrix}.$$

Then

$$A_1 = \begin{bmatrix} a_{11} & \mathbf{b}^\mathsf{T} \\ \mathbf{b} & B \end{bmatrix} = L_1 \begin{bmatrix} 1 & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & B - \frac{1}{a_{11}}\mathbf{b}\mathbf{b}^\mathsf{T} \end{bmatrix} L_1^\mathsf{T}.$$

Repeat this on the symmetric matrix $A_2 := B - \frac{1}{a_{11}}\mathbf{b}\mathbf{b}^\mathsf{T} \in \mathbb{R}^{(n-1)\times(n-1)}$.
Let $L_2, L_3, \dots, L_n$ be the matrices obtained in repeated steps. The matrix L is then

$$L = L_1 \cdot \begin{bmatrix} 1 & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & L_2 \end{bmatrix} \cdots \begin{bmatrix} I_{n-1} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & L_n \end{bmatrix}. \tag{1.2}$$

Chapter 2

Vector spaces and linear maps

2.1 Vector spaces

Problem 2.1 (solution 2.1)

Which subsets of the vector space $\mathbb{R}^{n\times n}$—all real $n \times n$ matrices—are vector subspaces? Determine the dimension of those that are.

(a) All matrices, which have 0 as the $(1,2)$-entry.

(b) All matrices, which have 1 as the $(1,2)$-entry.

(c) All matrices with integer entries, i.e. for $A = [a_{ij}]$ we have $a_{ij} \in \mathbb{Z}$.

(d) All upper-triangular matrices.

(e) All symmetric matrices; $A = A^\mathsf{T}$.

(f) All antisymmetric matrices; $A = -A^\mathsf{T}$.

(g) All invertible matrices; the subset $\mathrm{GL}(n,\mathbb{R}) \subseteq \mathbb{R}^{n\times n}$.

(h) All matrices with determinant 0, i.e. $\mathbb{R}^{n\times n} \setminus \mathrm{GL}(n,\mathbb{R})$.

(i) All nilpotent matrices, ie. matrices $N \in \mathbb{R}^{n\times n}$ such that $N^n = 0$.

(j) All upper-triangular nilpotent matrices. *(Hint: Which elements appear on the diagonal of an upper-triangular nilpotent matrix?)*

(k) All matrices with trace 0.

Problem 2.2 (solution 2.2)

Let F be the set of all Fibonacci sequences, i.e. sequences

$$(a_n)_{n=0}^{\infty} = (a_0, a_1, a_2, \ldots),$$

where a_0 and a_1 are arbitrary real numbers, and $a_n = a_{n-1} + a_{n-2}$ holds for all $n \geq 2$.

(a) Show that F is a vector space under operations

$$(a_n) + (b_n) := (a_n + b_n) \quad \text{and} \quad \alpha(a_n) := (\alpha a_n),$$

where $\alpha \in \mathbb{R}$.

(b) Find a basis for F and express the usual Fibbonacci sequence (the one with $a_0 = a_1 = 1$) in this basis.

Problem 2.3 (solution 2.3)

Equip the open interval $\mathbb{R}^+ = (0, \infty)$ with the operation $x \oplus y := xy$, and define $\alpha \odot x = x^\alpha$ for scalars $\alpha \in \mathbb{R}$.

(a) Show that $(\mathbb{R}^+, \oplus, \odot)$ is a vector space over $\mathbb{R}$.

(b) Find a basis for $(\mathbb{R}^+, \oplus, \odot)$ and determine its dimension.

Problem 2.4 (solution 2.4)

Let N be the matrix

$$N = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Show that the set of all real 2×2 matrices which commute with N, i.e.

$$U = \{A \in \mathbb{R}^{2\times 2} : AN = NA\},$$

is a vector subspace in $\mathbb{R}^{2\times 2}$. Find a basis for U and determine its dimension!

Problem 2.5 (solution 2.5)

(a) Is the set $U_1 = \{p(x) = ax + b : a \neq 0, a, b \in \mathbb{R}\}$ a vector subspace in the vector space of polynomials $\mathbb{R}_1[x]$?

(b) Is the set $U_2 = \{p(x) : p(0) = 0\}$ a vector subspace in the vector space of polynomials $\mathbb{R}_2[x]$?

(c) Is the set $U_3 = \{p(x) : p(0) = 1\}$ a vector subspace in the vector space of polynomials $\mathbb{R}_n[x]$?

(d) Is the set $U_4 = \{p(x) : p''(3) = 0\}$ a vector subspace in the vector space of polynomials $\mathbb{R}_n[x]$?

Problem 2.6

For a polynomial $p(x) = ax^3 + bx^2 + cx + d$ and a square matrix A denote $p(A) = aA^3 + bA^2 + cA + dI$. Let $A \in \mathbb{R}^{2\times 2}$ be the matrix

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

Let $U \subseteq \mathbb{R}_3[x]$ be the subset of those polynomials (of degree at most 3), for which $p(A) = 0$ (the zero matrix).

(a) Show that U is a vector subspace of $\mathbb{R}_3[x]$.

(b) Find a basis for U and determine $\dim U$.
(*Hint:* If $\Delta_A(x)$ is the characteristic polynomial of A, then $\Delta_A(A) = 0$.)

(c) Let $q(x) = x(x^2 - 2x - 3)$. Is the set of all 2×2 matrices X, for which $q(X) = 0$ holds, a vector subspace of $\mathbb{R}^{2\times 2}$? *Justify your answer!*

Problem 2.7 (solution 2.7)

Let $\mathbb{R}[x]$ be the set of all polynomials in the indeterminate x. (Hence, $\mathbb{R}[x]$ contains polynomials of arbitrary degrees!) Show that $\mathbb{R}[x]$ is a vector space for the usual addition of polynomials and multiplication of a scalar and a polynomial. Can you describe a basis for $\mathbb{R}[x]$? Can you find a basis for the subspace

$$W = \{p \in \mathbb{R}[x] : p(1) = p(-1) = 0\}?$$

Determine $\dim \mathbb{R}[x]$ and $\dim W$.

Problem 2.8 (solution 2.8)

Let $\mathbb{R}[[x]]$ be the set of all formal power series with real coefficients, the elements of $\mathbb{R}[[x]]$ are the (formal) sums

$$f(x) = \sum_{k=0}^{\infty} a_k x^k = a_0 + a_1 x + a_2 x^2 + \cdots,$$

which are added and multiplied by a scalar component wise. Check that $\mathbb{R}[[x]]$ too is a vector space. How is $\mathbb{R}[[x]]$ different from $\mathbb{R}[x]$? Can you find a basis for $\mathbb{R}[[x]]$?

Problem 2.9 (solution 2.9)

Let $V \subseteq C^\infty(0, 2\pi)$ be the set of all solutions to the differential equation

$$y'' + y = 0.$$

Show that V is a vector subspace of $C^\infty(0, 2\pi)$. Find its basis.

2.2 Linear maps

Problem 2.10 (solution 2.10)

A map $\tau\colon \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$ is given by

$$\tau(X) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} X + X \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

(a) Show that τ is a linear map.

(b) Find the matrix corresponding to τ with respect to the standard basis $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ of the vector space $\mathbb{R}^{2\times 2}$.

Problem 2.11 (solution 2.13)

For a polynomial $p(x) = ax^3 + bx^2 + cx + d$, and a square matrix A denote $p(A) = aA^3 + bA^2 + cA + dI$. Let $A \in \mathbb{R}^{2\times 2}$ be the matrix

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

(a) Show that the map given by

$$\phi\colon \mathbb{R}_3[x] \to \mathbb{R}^{2\times 2},\ \phi(p) = p(A)$$

is linear and determine the matrix corresponding to ϕ in the standard bases of spaces $\mathbb{R}_3[x]$ and $\mathbb{R}^{2\times 2}$.

(b) Find a basis for $\ker\phi$ and determine $\dim(\ker\phi)$. (*Hint:* If $\Delta_A(\lambda)$ is the characteristic polynomial of A, then $\Delta_A(A) = 0$.)

(c) Let $q(x) = x(x^2 - 2x - 3)$. Is the set of all 2×2 matrices X, for which $q(X) = 0$ holds, a vector subspace of $\mathbb{R}^{2\times 2}$?

Problem 2.12

(solution 2.12)

We are given vectors $\mathbf{a} = [1,1,0]^\mathsf{T}$, $\mathbf{b} = [1,0,1]^\mathsf{T}$, and $\mathbf{c} = [0,1,1]^\mathsf{T}$ in $\mathbb{R}^3$, and a linear map $\tau : \mathbb{R}^3 \to \mathbb{R}^3$ for which

$$\tau(\mathbf{a}) = \mathbf{a}, \tau(b) = \mathbf{a} + \mathbf{b}, \text{ and } \tau(\mathbf{c}) = \mathbf{a} + \mathbf{c}$$

holds.

(a) Show that $\{\mathbf{a},\mathbf{b},\mathbf{c}\}$ is a basis of $\mathbb{R}^3$.

(b) Find the matrix for τ in the basis $\mathcal{B} = \{\mathbf{a},\mathbf{b},\mathbf{c}\}$.

(c) Find the matrix for τ in the standard basis $\mathcal{S} = \{\mathbf{i},\mathbf{j},\mathbf{k}\}$.

(d) Where does the vector $[1,1,1]^\mathsf{T}$ get mapped by τ?

Problem 2.13

(solution 2.13)

Let $\mathbb{R}_3[x]$ be the vector space of polynomials p of degree at most 3.

(a) Check that the map $\phi\colon \mathbb{R}_3[x] \to \mathbb{R}^3$, $\phi(p) := [p(-1), p(0), p(1)]^\mathsf{T}$ is linear.

(b) Find a basis $\mathcal{B}_{\ker\phi}$ for the kernel $\ker\phi$ of the map ϕ.

(c) Find the matrix corresponding to ϕ in the basis $\{1, x, x^2, x^3\}$ of $\mathbb{R}_3[x]$ and the standard basis of $\mathbb{R}^3$.

Problem 2.14

(solution 2.14)

A map $\psi : \mathbb{R}_2[x] \to \mathbb{R}_2[x]$ is given by

$$(\psi(p))(x) = (xp(x+1))' - 2p(x).$$

Show that ψ is linear. Find its matrix in the basis $\{1, x, x^2\}$. Find its kernel and image.

Problem 2.15

(solution 2.15)

Let $\mathbf{a} = [1,1]^\mathsf{T}$. A map $\phi\colon \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ is given by

$$\phi(\mathbf{x}) = \mathbf{x}\,\mathbf{a}^\mathsf{T} = \mathbf{x}\,[1,1].$$

(a) Show that ϕ is linear.

(b) Find the matrix corresponding to ϕ in the standard bases of $\mathbb{R}^2$ and $\mathbb{R}^{2\times 2}$.

(c) Determine $\dim(\ker\phi)$ and $\dim(\operatorname{im}\phi)$.

(d) Find a basis of $\operatorname{im}\phi$.

Problem 2.16 (solution 2.16)

Assume that U and V are vector subspaces of a vector space W. Define sets:

$$U \times V := \{(u,v) : u \in U, v \in V\},$$
$$U + V := \{u + v : u \in U, v \in V\}, \text{ and}$$
$$U \cap V := \{w \in W : w \in U \text{ in } w \in V\}.$$

(a) Confirm that $U+V$ and $U \cap V$ are vector subspaces of W.

(b) 'Guess' the appropriate vector space structure on $U\times V$. Prove that $U\times V$ is actually a vector space in the guessed case! Determine $\dim(U\times V)$ from $\dim U$ and $\dim V$.

(c) Let a map $\phi\colon U\times V \to W$ be given by $\phi(u,v) = u - v$. Confirm that ϕ is linear. (If it turns out, that it is not, return to part (b).) Determine $\ker\phi$ and $\operatorname{im}\phi$.

(d) Show that the map $\psi\colon U\cap V \to \ker\phi$, $\psi(w) = (w,w)$ is a linear bijection, therefore $\dim(U\cap V) = \dim(\ker\phi)$.

(e) Conclude that $\dim U + \dim V = \dim(U+V) + \dim(U\cap V)$.

Chapter 3

Functions of several variables

3.1 Multiple integrals

Problem 3.1 (solution 3.1)

Let a vector–valued function $\mathbf{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $\mathbf{F}(\mathbf{r}) = \mathbf{F}(r,\varphi) = [x,y]^\top = \mathbf{x}$, where

$$\begin{aligned} x &= r\cos\varphi, \\ y &= r\sin\varphi. \end{aligned}$$

Find the Jacobi matrix $J_\mathbf{F} = \frac{\partial \mathbf{F}}{\partial \mathbf{r}}$ and the Jacobi determinant $\det(J_\mathbf{F})$ of $\mathbf{F}$.

Problem 3.2 (solution 3.2)

Let a vector–valued function $\mathbf{F}\colon \mathbb{R}^3 \to \mathbb{R}^3$ be given by $\mathbf{F}(\mathbf{r}) = \mathbf{F}(r,\varphi,z) = [x,y,z]^\top = \mathbf{x}$, where

$$\begin{aligned} x &= r\cos\varphi, \\ y &= r\sin\varphi, \\ z &= z. \end{aligned}$$

Find the Jacobi matrix $J_\mathbf{F} = \frac{\partial \mathbf{F}}{\partial \mathbf{r}}$ and the Jacobi determinant $\det(J_\mathbf{F})$ of $\mathbf{F}$.

Problem 3.3 (solution 3.3)

Let $\mathbf{F}\colon \mathbb{R}^3 \to \mathbb{R}^3$, $\mathbf{r} \mapsto \mathbf{x}$ be a vector–valued function given by $\mathbf{F}(r,\varphi,\theta) := [x,y,z]^\top$, where:

$$\begin{aligned} x &= r\cos\theta\cos\varphi, \\ y &= r\cos\theta\sin\varphi, \\ z &= r\sin\theta. \end{aligned}$$

Find the Jacobi matrix $J_\mathbf{F} = \frac{\partial \mathbf{F}}{\partial \mathbf{r}}$ and the Jacobi determinant $\det(J_\mathbf{F})$ of $\mathbf{F}$.

Problem 3.4

Let $R \geq 0$, and let $\mathbf{F}\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a vector–valued function given by

$$\mathbf{F}(r,\varphi,\theta) = \mathbf{F}\left([r,\varphi,\theta]^{\mathrm{T}}\right) = \begin{bmatrix} (R + r\cos\theta)\cos\varphi \\ (R + r\cos\theta)\sin\varphi \\ r\sin\theta \end{bmatrix}.$$

1. Find the Jacobi matrix $\mathbf{F}$; $J_{\mathbf{F}} = \frac{\partial \mathbf{F}}{\partial [r,\varphi,\theta]^{\mathrm{T}}}$ of $\mathbf{F}$.
2. Find the determinant of that Jacobi matrix; $\det(J_{\mathbf{F}})$.

Problem 3.5 (solution 3.5)

Evaluate double integrals below.

(a) $\displaystyle\iint_D (5 - x - y)\,dxdy$, where $D = [0,1]\times[0,1]$,

(b) $\displaystyle\iint_D \frac{y}{x+1}\,dxdy$, where D is given by $x \geq 0$, $y \geq x$ in $x^2 + y^2 \leq 2$,

(c) $\displaystyle\iint_D \frac{\sin x}{x}\,dxdy$, where D is the triangle given by $0 \leq y \leq x$, and $x \leq \pi$,

(d) $\displaystyle\iint_{\mathbb{R}^2} e^{-x^2-y^2}\,dxdy$, and use this to evaluate $\displaystyle\int_{-\infty}^{\infty} e^{-x^2}\,dx$.

Problem 3.6

Sketch the domain of integration and evaluate the integrals.

1. $\displaystyle\int_0^1 \left(\int_{-x}^{x} xe^y\,dy\right)dx,$
2. $\displaystyle\int_0^1 \left(\int_0^{y} \frac{y}{x+1}\,dx\right)dy + \int_1^{\sqrt{2}} \left(\int_0^{\sqrt{2-y^2}} \frac{y}{x+1}\,dx\right)dy.$

Problem 3.7 (solution 3.7)

What is the volume of the solid bounded by the paraboloid $z = 8 - x^2 - y^2$ and the plane $z = -1$?

Problem 3.8 (solution 3.8)

Find the coordinates of the center of mass of the quarter of a disk given by inequalities $x^2 + y^2 \leq R^2$, $x \geq 0$, $y \geq 0$ if the density at each point is equal to the distance from the origin, ie. $\rho(x,y) = \sqrt{x^2 + y^2}$.
Hint: the mass of a figure $D \subseteq \mathbb{R}^2$ is given by $m = \iint_D \rho(x,y)dxdy$, coordinates of the center of mass are $x^* = \frac{1}{m}\iint_D x\rho(x,y)dxdy$ and $y^* = \frac{1}{m}\iint_D y\rho(x,y)dxdy$. Use polar coordinates.

Problem 3.9 (solution 3.9)

Determine the mass and the coordinates of the center of mass of a homogeneous solid (ie. $\rho(x,y,z) = 1$) bounded by surfaces $z^2 = x^2+y^2$ and $x^2+y^2+z^2 = 4$, which lies in the half-space $z \geq 0$.
Hint: Use spherical coordinates:

$$\begin{aligned} x &= r\cos\theta\cos\varphi, \\ y &= r\cos\theta\sin\varphi, \\ z &= r\sin\theta, \end{aligned}$$

ie. a 'new variable' $\mathbf{F}(r,\varphi,\theta) = [x,y,z]^\mathsf{T}$ (for which $\det(J_\mathbf{F}) = r^2\cos\theta$ holds.)

Problem 3.10 (solution 3.10)

Determine the mass and the coordinates of the center of mass of a ball given by inequality $x^2+y^2+z^2 \leq 2z$ if the density at every point is equal to the distance from the origin.
Hint: Use spherical coordinates.

Problem 3.11 (solution 3.11)

A solid $D \subseteq \mathbb{R}^3$ is bounded by parabolic cylinders $z = 2-x^2$ and $z = y^2-2$. Determine the volume and mass of this solid if the density is given by $\rho(x,y,z) = y^2$.
Hint: Find the (orthogonal) projection of this solid onto the xy-plane, use cylindrical coordinates.

3.2 Local extrema of real multivariate functions

Problem 3.12 (solution 3.12)

Find and classify the stationary points of functions below.

(a) $f(x,y) = x^3 - 4x^2 + 2xy - y^2$
(b) $g(x,y) = xe^x + 2ye^y + 1$
(c) $h(x,y) = (1+e^y)\cos x - ye^y$
(d) $k(x,y,z) = x^3 + y^3 + 3z^2 - 3xyz$
(e) $r(x,y,z) = x^2 + y^2 + z^2 - 2xyz$
(f) $u(x,y) = 4 + x^3 + y^3 - 3xy$
(g) $v(x,y) = 3x^2y + y^3 - 3x^2 - 3y^2$

Problem 3.13 (solution 3.13)

Given $\mathbf{a},\mathbf{b} \in \mathbb{R}^n$ let $f(\mathbf{x}) = (\mathbf{x}^\mathsf{T}\mathbf{a})(\mathbf{x}^\mathsf{T}\mathbf{b})$.

(a) Evaluate $\frac{\partial f}{\partial \mathbf{x}}$ and $\frac{\partial^2 f}{\partial \mathbf{x}^2}$.
(b) Additionally assume that $\mathbf{a}$ and $\mathbf{b}$ are nonzero and orthogonal. What is the type of the lone stationary point of f in this case?

Problem 3.14 (solution 3.14)

Find the vector $\mathbf{x} \in \mathbb{R}^n$ for which the sum of squared distances from given vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{R}^n$ is the smallest possible.

3.3 Constrained extrema

Problem 3.15 (solution 3.15)

Find the points in the domain described by the inequality

$$4(x-1)^2 + y^2 \leq 16,$$

at which the largest and least values of the function

$$f(x,y) = 2x^2 + y^2$$

are attained.

Problem 3.16 (solution 3.16)

Let T be the triangle which is the intersection of the first octant and the plane given by $x + y + z = 5$. At which point on this triangle is the largest value of the function $g(x,y,z) = xy^2z^2$ attained?

Problem 3.17 (solution 3.17)

Find all points on the ellipse given by

$$x^2 - xy + y^2 = 3,$$

which are farthest from the origin.

Problem 3.18 (solution 3.18)

Which points on the curve given implicitly by

$$(x^2 + y^2)^2 = x^3 + y^3$$

are farthest from the origin?

Problem 3.19 (solution 3.19)

Find the largest and the least value of the function $f(x,y) = xy - y + x - 1$

(a) on the disk given by $x^2 + y^2 \leq 2$,

(b) on the half-disk given by $x^2 + y^2 \leq 2$ and $x \geq 0$.

Problem 3.20 (solution 3.20)

An ellipsoid is given by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

A box with edges parallel to x, y, and z axes is inscribed inside this ellipsoid.

(a) What is the largest possible volume of the inscribed box?

(b) What is the largest possible surface area of the inscribed box?

Problem 3.21 (solution 3.21)

We are given an ℓ meters long thin rod. We cut it into 12 shorter rods, from which a frame of a box can be assembled.

(a) How long should those shorter rods be if the box is to have largest possible volume?

(b) Same question as above with an additional restriction – the base rectangle should have area equal to A.

Problem 3.22 (solution 3.22)

We wish to assemble a frame of a triangular prism (equilateral base triangle) from an ℓ metres long thin rod.

(a) What should be the length of the pieces we cut our rod into, for the prism to have largest possible volume?

(b) What should be the length of the pieces we cut our rod into, for the prism to have largest possible surface area?

Problem 3.23 (solution 3.23)

Let $\mathbf{a} \in \mathbb{R}^n$ and let $d \geq 0$ be a real number.

(a) Find the largest and least value of the expression $\mathbf{a}^\mathsf{T}\mathbf{x}$ for a vector $\mathbf{x} \in \mathbb{R}^n$ with prescribed length $\|\mathbf{x}\| = d$.

(b) Explain the solution.

Problem 3.24 (solution 3.24)

Assume $A \in \mathbb{R}^{n\times n}$, and let d be a positive real number.

(a) Find the largest and least value of $f(\mathbf{x}) = \mathbf{x}^\mathsf{T} A\mathbf{x}$ with constraint $\|\mathbf{x}\| = d$.

(b) Find the largest and least value of $f(\mathbf{x}) = \|\mathbf{x}\|^2$ with constraint $\mathbf{x}^\mathsf{T} A\mathbf{x} = d^2$ if A is symmetric positive definite.

Problem 3.25 (solution 3.25)

Let $A \in \mathbb{R}^{n\times n}$, $\mathbf{b} \in \mathbb{R}^n$, and $d > 0$ a real number.

(a) Minimize $f(\mathbf{x}) = \|\mathbf{x}\|^2$ with respect to $\|\mathbf{x} - \mathbf{p}\| \leq d$.

(b) Minimize $f(\mathbf{x}) = \|\mathbf{x}\|^2$ with respect to $A\mathbf{x} = \mathbf{b}$.

(c) Minimize $f(\mathbf{x}) = \|\mathbf{x}\|^2$ with respect to $\|\mathbf{x} - \mathbf{p}\| \leq d$ and $A\mathbf{x} = \mathbf{b}$.

Chapter 4

Solutions

Solution to problem 1.1, page 3: First we compute the the characteric polynomial of A.

$$\begin{aligned}
p_A(\lambda) &= \det(A-\lambda I) = \begin{vmatrix} -\lambda & 2 & 2 \\ 3 & 1-\lambda & -3 \\ -1 & -1 & 3-\lambda \end{vmatrix} \\
&= \begin{vmatrix} -\lambda & 2 & 2-\lambda \\ 3 & 1-\lambda & 0 \\ -1 & -1 & 2-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 1-\lambda & 3 & 0 \\ 3 & 1-\lambda & 0 \\ -1 & -1 & 2-\lambda \end{vmatrix} \\
&= (2-\lambda)\begin{vmatrix} 1-\lambda & 3 \\ 3 & 1-\lambda \end{vmatrix} \\
&= (2-\lambda)((1-\lambda)^2-9) \\
&= (2-\lambda)(4-\lambda)(-2-\lambda)
\end{aligned}$$

The eigenvalues are the roots of the characteristic polynomial (indexed in decreasing order)

$$\lambda_1 = 4,\ \lambda_2 = 2,\ \lambda_3 = -2$$

To find the corresponding eigenvectors we determine the null space

$$N(A-\lambda I) = \{\mathbf{v} \in \mathbb{R}^3, (A-\lambda I)\mathbf{v} = 0\}$$

for each eigenvalue λ, which can be done using Gaussian elimination.

1. For $N(A-\lambda_1 I)$ we compute

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

   Our homogenous system of equations for the (x, y, z) coordinates are thus simplified to

$$\begin{aligned} x &= 0 \\ y+z &= 0 \end{aligned}$$

We can take z to be the free variable and choosing $z = 1$ gives an eigenvector for $\lambda_1 = 4$.

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

2. For $N(A - \lambda_2 I)$ we have

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

The equations are $x - z = 0$ and $y = 0$. Choosing $z = 1$ gives

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

3. For $N(A - \lambda_3 I)$ we have

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

The equations are $x + y = 0, z = 0$. This time we take y for the free variable and the choice $y = 1$ gives

$$\mathbf{v}_3 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

Solution to problem 1.2, page 3: We know that it will be possible to find an orthogonal basis for $\mathbb{R}^4$ consisting of eigenvectors of H because H is a symmetric matrix ($H^\mathsf{T} = H$). Computing the characteristic polynomial for H we get

$$\begin{aligned}
\det(H - \lambda I) &= \begin{vmatrix} 1-\lambda & 1 & 1 & 0 \\ 1 & -\lambda & 0 & 1 \\ 1 & 0 & -\lambda & -1 \\ 0 & 1 & -1 & 1-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 1-\lambda & 1 & 1 & 0 \\ 2 & -\lambda & 0 & 1 \\ 0 & 0 & -\lambda & -1 \\ 1-\lambda & 1 & -1 & 1-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 1-\lambda & 1 & 1 & 0 \\ 2 & -\lambda & 0 & 1 \\ 0 & 0 & -\lambda & -1 \\ 0 & 0 & -2 & 1-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 1-\lambda & 1 \\ 2 & -\lambda \end{vmatrix} \cdot \begin{vmatrix} -\lambda & -1 \\ -2 & 1-\lambda \end{vmatrix} \\
&= (-(1-\lambda)\lambda - 2)^2 \\
&= (2-\lambda)^2(-1-\lambda)^2
\end{aligned}$$

In addition to the usual rules for computing determinants (addition of multiples of rows/columns to other rows/columns) we used the following rule for block-upper-diagonal determinants:

$$\begin{vmatrix} A & C \\ 0 & B \end{vmatrix} = |A| \cdot |B|$$

where A, B, C and 0 are matrices of appropriate dimensions.

Therefore obtain two distinct eigenvalues $\lambda_{1,2} = 2$ and $\lambda_{3,4} = -1$ which are both double roots of the characteristic polynomial. Since H is symmetric, we know we will have two-dimensional eigenspaces for each of the eigenvalues (for general matrices this is not necessarily the case).

Solution to problem 1.3, page 3:

(a) Let us first assume A is diagonalizable, so that we have

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

where $D = diag(\lambda_1, \ldots, \lambda_n)$ is diagonal matrix that contains the eigenvalues of A along the diagonal. The powers A^k of the matrix A can then be expressed as

$$A^k = PDP^{-1}PDP^{-1} \cdots PDP^{-1} = PD^kP^{-1}$$

where $D^k = diag(\lambda_1^k, \ldots, \lambda_n^k)$ contains the powers of the eigenvalues along the diagonal. We use this expression in the power series for e^A.

$$\begin{aligned} e^A &= \sum_{k=0}^{\infty} \frac{1}{k!} A^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} PD^kP^{-1} \\ &= P\left(\sum_{k=0}^{\infty} \frac{1}{k!} D^k \right) P^{-1} \\ &= P \cdot diag(\sum_{k=0}^{\infty} \frac{1}{k!} \lambda_1^k, \ldots, \sum_{k=0}^{\infty} \frac{1}{k!} \lambda_n^k) \cdot P^{-1} \\ &= P \cdot diag(e^{\lambda_1}, \ldots, e_n^{\lambda}) \cdot P^{-1} \end{aligned}$$

In other words, the matrix e^A can be diagonalised in the same basis as A, and its eigenvalues are simply $e^{\lambda_1}, \ldots, e^{\lambda_n}$. Using elementary properties of determinants, the determinant of e^A can then be expressed as

$$\begin{aligned} \det(e^A) &= \det(P \cdot diag(e^{\lambda_1}, \ldots, e_n^{\lambda}) \cdot P^{-1}) \\ &= \det(P) \cdot \det(diag(e^{\lambda_1}, \ldots, e_n^{\lambda})) \cdot \det(P^{-1}) \\ &= \det(diag(e^{\lambda_1}, \ldots, e_n^{\lambda})) \\ &= e^{\lambda_1} \cdots e_n^{\lambda} = e^{\lambda_1 + \ldots + \lambda_n} \end{aligned}$$

We know that $\mathrm{tr}(A)$ equals the sum of the eigenvalues of A which completes the proof for diagonalizable A.

For the general case we can use the Schur decomposition

$$A = UTU^*$$

where T is an upper-triangular matrix and U is an unitary matrix ($U^*U = UU^* = I$). We know that the upper-triangular matrix in the Schur decomposition of A also contains the eigenvalues of A on its diagonal. The proof then follows the same pattern as in the diagonalizable case, with the difference being that we work with upper-triangular matrices instead of diagonal matrices. However these upper-triangular matrices that appear contain the same diagonal elements as the diagonal matrices above. Since the matrix elements above the diagonal do not affect the determinant or the trace, this means the result is the same in the end.

(b) Using the power series for e^A we can directly see that

$$\left(e^A\right)^\mathsf{T} = e^{A^\mathsf{T}}$$

For matrices A and B that commute (meaning $AB = BA$) it is also possible to show that

$$e^A \cdot e^B = e^{A+B}$$

(This is identity is *not* valid for matrices that do not commute!). For an antisymmetric matrix A we then have

$$\left(e^A\right)^\mathsf{T} \cdot e^A = e^{A^\mathsf{T}} \cdot e^A = e^{-A} \cdot e^A = e^{-A+A} = e^0 = I$$

since A and $-A$ commute (the zero above denotes the zero matrix), proving e^A is an orthogonal matrix. If $A^\mathsf{T} = -A$, then all elements on the diagonal of A must be equal to zero, implying that $\mathrm{tr}(A) = 0$. Using the identity from (a) we then have

$$\det(A) = e^{\mathrm{tr}(A)} = e^0 = I$$

Solution to problem 1.4, page 4:

(a) By performing Gaussian elimination

$$A = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

or by simply looking at the matrix A it is clear that the column space $C(A)$ is spanned by the first two column vectors. We denote these two vectors by

$$\mathbf{u} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

so we can write $C(A) = Lin\{\mathbf{u},\mathbf{v}\}$ and conclude that the dimension of $C(A)$ equals 2.

This implies that $\dim(N(A)) = n-2$, which can also be seen from the row-reduced form of A above. To determine a basis for the null space $N(A)$, i.e. the subspace of the solutions to the equation $A\mathbf{w} = \mathbf{0}$, we need to find $n-2$ linearly independent solutions to the system of equations

$$\begin{aligned} x_1 &= 1 \\ x_2 + \ldots + x_n &= 0 \end{aligned}$$

where the variables are the coordinates of the unknown vector $\mathbf{w} = [x_1,\ldots,x_n]^{\mathsf{T}}$. This system of equations can be obtained directly from the result of Gaussian elimination above or by noticing that A is a symmetric matrix: Since we know that $C(A)^{\perp} = N(A^{\mathsf{T}}) = N(A)$, we can conclude that $\mathbf{w} \in N(A)$ is equivalent to the condition that $\mathbf{w}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$. Writing these conditions with equations, i.e. $\mathbf{w}\cdot\mathbf{u} = 0$ and $\mathbf{w}\cdot\mathbf{v} = 0$, using coordinates produces the same equations as above.

The 'free' variables of this system are $x_3,\ldots,x_n$ which means we can choose $n-2$ linearly independent solutions by setting $x_k = 1$, $x_i = 0$ for $i \neq k$ for each choice of $k = 3,\ldots,n$. The solutions can then be written as

$$\mathbf{w}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{w}_4 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \ldots, \mathbf{w}_{n-1} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}, \mathbf{w}_n = \begin{bmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

and hence one choice of a basis for $N(A)$ is the set of vectors $\{\mathbf{w}_3,\ldots,\mathbf{w}_n\}$.

(b) The $N(A)$ subspace is by definition the eigenspace of A for the eigenvalue $\lambda = 0$ (at least if $N(A)$ is non-trivial). We can therefore immediately conclude that we have a $n-2$ dimensional eigenspace for eigenvalue $\lambda = 0$ spanned by the vectors $\{\mathbf{w}_3,\ldots,\mathbf{w}_n\}$ defined above. Since A is a symmetric matrix, we know that the algebraic multiplicity of each eigenvalue is equal to the dimension of corresponding eigenspace. This means we have (algebraicly) two non-zero eigenvalues yet to be determined.

It is possible to obtain the remaining two eigenvalues as usual by computing the characteristic polynomial directly and then solving the corresponding systems of equations to obtain the eigenvectors. A better alternative is use the fact that A is a symmetric matrix.

Since we know that the eigenspaces of a symmetric matrix for distinct eigenvalues are mutually orthogonal, and that the orthogonal complement of the eigenspace for $\lambda = 0$ is $N(A)^{\perp} = C(A^{\mathsf{T}}) = C(A) = Lin\{\mathbf{u},\mathbf{v}\}$, we can deduce that it must be possible to express the remaining two eigenvectors as linear combinations of the column vectors $\mathbf{u}$ and $\mathbf{v}$.

It therefore makes sense to see how the matrix A acts on the vectors $\mathbf{u}$

and $\mathbf{v}$. Direct computation produces the equations

$$\begin{aligned} A\mathbf{u} &= (n-1)\mathbf{v} \\ A\mathbf{v} &= \mathbf{u} \end{aligned} \tag{1.1}$$

There are now at least two ways to proceed from here.

1. Since we know that any eigenvector $\mathbf{w}$ of A for any non-zero eigenvalue λ can be written as a linear combination $\mathbf{w} = x\mathbf{u} + y\mathbf{v}$, we can write the eigenvector equation $A\mathbf{w} = \lambda\mathbf{w}$ using (1.1) as follows

   $$A\mathbf{w} = A(x\mathbf{u} + y\mathbf{v}) = xA\mathbf{u} + yA\mathbf{v} = x(n-1)\mathbf{v} + y\mathbf{u} = \lambda(x\mathbf{u} + y\mathbf{v})$$

   The last equality can be rewritten as

   $$(\lambda x - y)\mathbf{u} + (x(n-1) - \lambda y)\mathbf{v} = \mathbf{0}$$

   Since $\mathbf{u}$ and $\mathbf{v}$ are linearly independent this implies that both coefficients above must equal 0, which yields a system of equations

   $$\begin{aligned} \lambda x - y &= 0 \\ x(n-1) - \lambda y &= 0 \end{aligned}$$

   By expressing $y = \lambda x$ from the first equation and plugging into the second equation we get $\lambda^2 = n-1$ or $\lambda_{1,2} = \pm\sqrt{n-1}$. (Of course this system also has an obvious solution for $x = y = 0$, but this is not a valid solution in this case because eigenvectors must be non-zero.) To obtain the eigenvectors for $\lambda_{1,2}$ we have freedom to choose any non-zero value for (say) x (because know that any non-zero multiple of an eigenvector is also an eigenvector for the same eigenvalue). We choose $x = 1$ and hence $y = \lambda$ for each eigenvalue $\lambda_{1,2}$ to get

   $$\mathbf{w}_1 = x\mathbf{u} + y\mathbf{v} = \mathbf{u} + \sqrt{n-1}\mathbf{v} = \begin{bmatrix} \sqrt{n-1} \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

   and

   $$\mathbf{w}_2 = \mathbf{u} - \sqrt{n-1}\mathbf{v} = \begin{bmatrix} -\sqrt{n-1} \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

   to get the remaining eigenvectors.

2. Another way is to use a little theory of linear maps (that are the subject of subsequent chapters). According to (1.1), we can view multiplication by A (restricted to 2-dimensional subspace) $C(A)$) as a linear map that $C(A)$ to $C(A)$. Using (1.1) we can represent this map in the $\{\mathbf{u}, \mathbf{v}\}$ basis by a 2×2 matrix, which we denote by B:

   $$B = \begin{bmatrix} 0 & 1 \\ n-1 & 0 \end{bmatrix}$$

Theory tells us that the eigenvalues of a linear map do not depend on the choice of basis for the underlying vector space, which means that the eigenvectors of B must be same as the restriction of A to $C(A)$. The characteristic polynomial for B is

$$\det(B-\lambda I) = \lambda^2 - (n-1) = 0$$

which yields $\lambda_{1,2} = \pm\sqrt{n-1}$. The the basis of he eigenspaces $N(B-\lambda_{1,2}I)$ (expressed with regard to the $\{\mathbf{u},\mathbf{v}\}$) can be computed as usual by Gaussian elimination.

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ \sqrt{n-1} \end{bmatrix}, \ \mathbf{w}_2 = \begin{bmatrix} 1 \\ -\sqrt{n-1} \end{bmatrix}$$

Since the coordinates of these vectors represent coefficients in the $\{\mathbf{u},\mathbf{v}\}$ of $C(A)$, we can express them as

$$\mathbf{w}_1 = \mathbf{u} + \sqrt{n-1}\mathbf{v}, \ \mathbf{w}_2 = \mathbf{u} + \sqrt{n-1}\mathbf{v}$$

Solution to problem 1.5, page 4:

The simplest way to determine the remaining eigenvalues of A (apart from the double eigenvalue 3) is to notice what happens when we add and subtract the equations $A\mathbf{v}_1 = \mathbf{v}_2$ and $A\mathbf{v}_2 = \mathbf{v}_1$.

By adding them we get

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

and by subtracting them we get

$$A(\mathbf{v}_1 - \mathbf{v}_2) = -(\mathbf{v}_1 - \mathbf{v}_2)$$

From these two equations we can directly read two eigenvectors

$$\mathbf{u}_1 = \mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{for eigenvalue } \lambda_1 = 1$$

and

$$\mathbf{u}_2 = \mathbf{v}_1 - \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix} \quad \text{for eigenvalue } \lambda_2 = -1$$

An alternative way to determine these two eigenvectors would be similar to that in Exercise 1.4 (b): we could write a 2×2 matrix that represents multiplication by A on the subspace spanned by $\{\mathbf{v}_1, \mathbf{v}_2\}$ and compute $\lambda_1, \lambda_2, \mathbf{u}_1, \mathbf{u}_2$ using the usual procedure.

We also notice that the computed eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal, which is a necessary condition for A to be a symmetric matrix. If the computed $\mathbf{u}_1$ and $\mathbf{u}_2$ were not orthogonal we could conclude that a symmetric matrix A with the required properties does not exist.

Now that we have all four eigenvalues $\lambda_1 = 1, \lambda_2 = -1$ (together with their eigenvectors) and $\lambda_{3,4} = 3$, we only need to determine the corresponding eigenspace for $\lambda_{3,4} = 3$.

The requirement that A must be a symmetric matrix implies that the eigenspace $N(A-3I)$ must be 2-dimensional and it must be orthogonal to the eigenspaces for λ_1 and λ_2. This means it is enough to find a basis for the orthogonal complement to $Lin\{\mathbf{u}_1, \mathbf{u}_2\}$ which is also 2-dimensional. The condition $\mathbf{x} \in Lin\{\mathbf{u}_1, \mathbf{u}_2\}^\perp$ can be described by equations $\mathbf{u}_1 \cdot \mathbf{x} = 0$ and $\mathbf{u}_2 \cdot \mathbf{x} = 0$, i.e. a system of equations

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 0 \\ -x_1 + x_2 + x_3 - x_4 &= 0 \end{aligned}$$

for the coordinates $\mathbf{x} = [x_1, x_2, x_3, x_4]^\mathsf{T}$.

Another way to obtain this system of equations is to define the matrix

$$U = [\mathbf{u}_1, \mathbf{u}_2] = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and note that

$$Lin\{\mathbf{u}_1, \mathbf{u}_2\}^\perp = C(U)^\perp = N(U^\mathsf{T})$$

In any case, we can perform Gaussian elimination on the matrix U^T to obtain

$$U^\mathsf{T} \sim \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

or on the system above to directly to obtain the conditions

$$\begin{aligned} x_1 + x_4 &= 0 \\ x_2 + x_3 &= 0 \end{aligned}$$

By choosing the appropriate values for the 'free' variables x_3 and x_4 we can choose the following basis for the eigenspace $N(A-3I) = N(U^\mathsf{T})$

$$\mathbf{u}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{u}_4 = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

By finding all the eigenvalues and eigenvectors we effectively have a diagonalisation of the matrix A. It remains to define the matrices

$$D = diag(1, -1, 3, 3) \text{ and } P = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4]$$

and compute $A = PDP^{-1}$. If we are computing this product by hand, it may be preferable to choose an orthonormal basis of eigenvectors (in order to avoid computing the inverse P^{-1}) instead of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$. Luckily, all these vectors are already orthogonal, so we only need to normalize them. We can define therefore define an orthonormal basis of eigenvectors by

$$\mathbf{q}_1 = \frac{1}{2}\mathbf{u}_1, \mathbf{q}_2 = \frac{1}{2}\mathbf{u}_2, \mathbf{q}_3 = \frac{1}{\sqrt{2}}\mathbf{u}_3, \mathbf{q}_4 = \frac{1}{\sqrt{2}}\mathbf{u}_4,$$

define the matrix $Q = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4]$ and compute

$$A = QDQ^\top = \frac{1}{2}\begin{bmatrix} 3 & 1 & 1 & -3 \\ 1 & 3 & -3 & 1 \\ 1 & -3 & 3 & 1 \\ -3 & 1 & 1 & 3 \end{bmatrix}$$

Solution to problem 1.6, page 4: In principle, a Schur decomposition for a matrix $A \in \mathbb{R}^{n\times n}$ can be computed by the following algorithm.

We first find an eigenvalue λ_1 for A and a corresponding normalized eigenvector $\mathbf{q}_1$, so we have $A\mathbf{q}_1 = \lambda_1\mathbf{q}_1$ with $\mathbf{q}_1^\top\mathbf{q}_1 = 1$. Then we find an ONB $\{\mathbf{q}_2, \dots, \mathbf{q}_n\}$ for the orthogonal complement of $\mathbf{q}_1$. In other words we form a an orthogonal matrix

$$Q_1 = \left[\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\right]$$

meaning $Q_1^\top Q_1 = I$, or, equivalently, $\mathbf{q}_i^\top\mathbf{q}_i = 1$ and $\mathbf{q}_i^\top\mathbf{q}_j = 0$ for $i \neq j$, $i,j = 1,\dots,n$. Then we compute the matrix

$$\begin{aligned} T_1 &:= Q_1^\top A Q_1 \\ &= \begin{bmatrix} \mathbf{q}_1^\top \\ \mathbf{q}_2^\top \\ \vdots \\ \mathbf{q}_n^\top \end{bmatrix} A \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \dots & \mathbf{q}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{q}_1^\top \\ \mathbf{q}_2^\top \\ \vdots \\ \mathbf{q}_n^\top \end{bmatrix} \begin{bmatrix} \lambda_1\mathbf{q}_1 & A\mathbf{q}_2 & \dots & A\mathbf{q}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\mathbf{q}_1^\top\mathbf{q}_1 & \mathbf{q}_1^\top A\mathbf{q}_2 & \dots & \mathbf{q}_1^\top A\mathbf{q}_n \\ \lambda_1\mathbf{q}_2^\top\mathbf{q}_1 & \mathbf{q}_2^\top A\mathbf{q}_2 & \dots & \mathbf{q}_2^\top A\mathbf{q}_n \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1\mathbf{q}_n^\top\mathbf{q}_1 & \mathbf{q}_n^\top A\mathbf{q}_2 & \dots & \mathbf{q}_n^\top A\mathbf{q}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & \mathbf{b}^\top \\ \mathbf{0} & A_2 \end{bmatrix} \end{aligned}$$

where the vector $\mathbf{b}$ and matrix A_2 are simply the results of the computation. This basically yields the first column and first row of upper triangular matrix T in the Schur decomposition $A = QTQ^\top$.

We can then repeat the procedure on the A_2 block of the matrix T_1 to obtain $(n-1)\times(n-1)$ matrices Q_2 and T_2 and so on. The end result is the upper-triangular matrix from the Schur decomposition together with a sequence $Q_1, Q_2, \dots, Q_{n-1}$ of orthognal matrices of decreasing size. The orthognal martix Q from the Schur decomposition can then be computed by

$$Q = Q_1 \begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & Q_2 \end{bmatrix} \cdots \begin{bmatrix} I_{n-1} & \mathbf{0}^\top \\ \mathbf{0} & Q_{n-1} \end{bmatrix}$$

It should be noted that we have considerable freedom during the computing of the Schur decomposition using the described algorithm, from the choice of the 'first' eiqenvalue to the choice of ONB for the orthogonal complement of the chosen eigenvector at each step of the algorithm. When computing by hand it is therefore worth it to use this freedom at each step with an eye towards keeping the subsequent computations as simple as possible.

We will find the Schur decomposition of matrix B first, since it is the easier example. The characteristic polynomial is

$$\det(B-\lambda I) = \begin{vmatrix} 2-\lambda & -1 & 0 \\ 0 & 1-\lambda & 0 \\ -\sqrt{2} & -\sqrt{2} & 2-\lambda \end{vmatrix} = (2-\lambda)\begin{vmatrix} 2-\lambda & 0 \\ 0 & 1-\lambda \end{vmatrix} = (2-\lambda)^2(1-\lambda)$$

We choose the 'first' eigenvalue to be $\lambda = 1$ since we easily notice that the corresponding eigenvector is simply $\mathbf{q}_1 = [0,0,1]^\mathsf{T}$. Using Gaussian elimination we could also verify that $\mathbf{q}_1$ is also the only eigenvector for the double eigenvalue$\lambda = 2$ which means that B is not diagonalizable, but a Schur decomposition still exists.

Clearly, we have many choices for the orthogonal matrix Q_1 which should contain $\mathbf{q}_1$ in the first column. A good choice is

$$Q_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

We now compute

$$\begin{aligned} T_1 &= Q_1^\mathsf{T} B Q_1 \\ &= \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 0 \\ -\sqrt{2} & -\sqrt{2} & 2 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 2 & -1 \\ 0 & 0 & 1 \\ 2 & -\sqrt{2} & -\sqrt{2} \end{bmatrix} \\ &= \begin{bmatrix} 2 & -\sqrt{2} & -\sqrt{2} \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

The matrix multiplication in this case is easy since Q_1 happens to be a permutation matrix: multiplication by Q_1 from the right just permutes column vectors and multiplication from the left permutes row vectors.

We notice that the result T_1 already happens to be an upper-triangular matrix so no further steps are needed. The Schur decomposition of B is simply $B = Q_1 T_1 Q_1^\mathsf{T}$.

Of course, a less fortunate choice of Q_1 would require more computation. For instance, a sensible choice for Q_1 also seems to be

$$Q_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

This choice then results in

$$T_1 = \begin{bmatrix} 2 & -\sqrt{2} & -\sqrt{2} \\ 0 & 0 & 1 \\ 0 & 2 & -1 \end{bmatrix}$$

and the matrix

$$B_2 = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}$$

As before, we can notice the eigenvalue $\lambda = 2$ (we also know this must be an eigenvalue since the matrix $Q_1^\mathsf{T} B Q_1$ has the same eigenvalues as B) with eigenvector $\mathbf{q}_2 = [0, 1]^\mathsf{T}$. The (almost) only choice for Q_2 is then

$$Q_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and we get

$$T_2 = Q_2^\mathsf{T} B_2 Q_2 = \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix}$$

(Q_2 is a permutation matrix that just swaps the first and second columns/rows of matrices.)

The final result is

$$T = \begin{bmatrix} 2 & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & T_2 \end{bmatrix} = \begin{bmatrix} 2 & -\sqrt{2} & -\sqrt{2} \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

for the upper-triangular matrix of the decomposition and

$$Q = Q_1 \begin{bmatrix} 1 & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & Q_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

for the orthogonal matrix, which is the same result as before.

A somewhat more involved example is matrix A. The characteristic polynomial is

$$\begin{aligned}
\det(A - \lambda I) &= \begin{vmatrix} 6-\lambda & -1 & 1 \\ 4 & 3-\lambda & 1 \\ 2 & 2 & 3-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 6-\lambda & -1 & 1 \\ \lambda-2 & 4-\lambda & 0 \\ 2 & 2 & 3-\lambda \end{vmatrix} \\
&= \begin{vmatrix} \lambda-2 & 4-\lambda \\ 2 & 2 \end{vmatrix} + (3-\lambda) \begin{vmatrix} 6-\lambda & -1 \\ \lambda-2 & 4-\lambda \end{vmatrix} \\
&= (2(\lambda-2) - 2(4-\lambda)) + (3-\lambda)((6-\lambda)(4-\lambda) + (\lambda-2)) \\
&= -4(3-\lambda) + (3-\lambda)(\lambda^2 - 9\lambda + 22) \\
&= (3-\lambda)(\lambda^2 - 9\lambda + 18) \\
&= (3-\lambda)^2(6-\lambda)
\end{aligned}$$

Let us find an eigenvector for $\lambda = 6$. Using Gaussian elimination we get

$$A - 6I \sim \begin{bmatrix} 0 & -1 & 1 \\ 4 & -3 & 1 \\ 2 & 2 & -3 \end{bmatrix} \sim \begin{bmatrix} 2 & 2 & -3 \\ 0 & -1 & 1 \\ 4 & -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 2 & 2 & -3 \\ 0 & -1 & 1 \\ 0 & -7 & 7 \end{bmatrix} \sim \begin{bmatrix} 2 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

The equations for the coordinates of an eigenvector $\mathbf{v}_1 = [x_1, x_2, x_3]^\mathsf{T}$ for $\lambda = 6$ therefore reduce to

$$\begin{aligned} 2x_1 - x_3 &= 0 \\ -x_2 + x_3 &= 0 \end{aligned}$$

If we choose $x_3 = 2$ for the value of the 'free' variable we get the eigenvector $\mathbf{v}_1 = [1, 2, 2]^\mathsf{T}$. Similarly, we can find that the only eigenvectors for the eigenvalue $\lambda = 3$ are multiples of $\mathbf{v}_2 = [-1, 1, 4]^\mathsf{T}$.

However, the eigenvector $\mathbf{v}_1$ seems 'nicer' than $\mathbf{v}_2$. For one, the length $|\mathbf{v}_1| = 3$ is an integer which means we don't need to deal with square-roots when normalizing. Also, it is possible to find two mutually orthogonal vectors to $\mathbf{v}_1$ simply by cleverly permuting and changing signs of its coordinates, which also ensures they all have the same length.

After some guesswork, a sensible choice for Q_1 seems to be

$$Q_1 = \frac{1}{3}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix}$$

The result for T_1 is

$$T_1 = Q_1^\mathsf{T} A Q_1 = Q_1^\mathsf{T}\begin{bmatrix} 2 & 3 & 5 \\ 4 & 3 & 1 \\ 4 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 3 & 3 \\ 0 & 3 & 3 \\ 0 & 0 & 3 \end{bmatrix}$$

Since T_1 happens to be an upper-triangular matrix we can terminate the algorithm and write the Schur decomposition as $A = Q_1 T_1 Q_1^\mathsf{T}$.

With a little less luck with our choice for Q_1, for instance if we had chosen

$$Q_1 = \frac{1}{3}\begin{bmatrix} 1 & 2 & 2 \\ 2 & -2 & 1 \\ 2 & 1 & -2 \end{bmatrix}$$

the result for T_1 would be

$$T_1 = \begin{bmatrix} 6 & 3 & 3 \\ 0 & 0 & 3 \\ 0 & 3 & 3 \end{bmatrix}$$

The algorithm would then require one more step. But even in this case we can notice that T_1 can be transformed into an upper-triangular matrix by the same permutation matrix that transposes the second and third columns and rows as in the example for the B matrix above.

Solution to problem 1.7, page 4:

1. By the definition $\|A\|_F = \sqrt{\operatorname{tr}(A^\mathsf{T} A)}$ and assumption $U^\mathsf{T} U = I$ we have

$$\|UA\|_\mathrm{F}^2 = \operatorname{tr}((UA)^\mathsf{T} UA) = \operatorname{tr}(A^\mathsf{T} U^\mathsf{T} UA) = \operatorname{tr}(A^\mathsf{T} A) = \|A\|_F^2$$

2. Here, we also need a basic property of the trace operation $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

$$\|AV\|_\mathrm{F}^2 = \operatorname{tr}((AV)^\mathsf{T} AV) = \operatorname{tr}(V^\mathsf{T} A^\mathsf{T} V) = \operatorname{tr}(A^\mathsf{T} AVV^\mathsf{T}) = \operatorname{tr}(A^\mathsf{T} A) = \|A\|_F^2$$

3. This can also be proved directly by definition, or simply by combining the previous two equalities

$$\|UAV\|_F = \|AV\|_F = \|A\|_F$$

Solution to problem 1.8, page 5:

1. Define the function $f(x) := \|Ax+B\|_F^2$ for $x \in \mathbb{R}$. Clearly, $f(x) \geq 0$ for all $x \in \mathbb{R}$. Expanding according to the definitions and the properties of the inner product we have

$$\begin{aligned} f(x) &= \|Ax+B\|_F^2 \\ &= \langle Ax+B, Ax+B\rangle_F \\ &= \langle Ax, Ax\rangle_F + \langle Ax, B\rangle_F + \langle B, Ax\rangle_F + \langle B, B\rangle_F \\ &= \langle A, A\rangle_F \cdot x^2 + 2\langle A, B\rangle_F \cdot x + \langle B, B\rangle_F \end{aligned}$$

This shows $f(x)$ is a quadratic function of x. A quadratic function $f(x)$ is non-negative for all $x \in \mathbb{R}$ if and only if its discriminant $D = b^2 - 4ac$ is non-positive, $D \leq 0$. In our case the discriminant equals

$$D = 4(\langle A, B\rangle_F)^2 - 4\langle A, A\rangle_F \langle B, B\rangle_F = 4(\langle A, B\rangle_F)^2 - 4\|A\|_F^2 \|B\|_F^2$$

The condition $D \leq 0$ then gives $(\langle A, B\rangle_F)^2 \leq \|A\|_F^2\|B\|_F^2$ which implies $|\langle A, B\rangle_F| \leq \|A\|_F\|B\|_F$.

We also mention that the Cauchy–Schwarz inequality holds not only for the Frobenius inner product but for general inner products on a vector spaces. The only properties of the inner product needed to prove the Cauchy-Schwarz inequality were $\langle A, A\rangle \geq 0$, $\langle A, B\rangle = \langle B, A\rangle$ and $\langle xA, B\rangle = x\langle A, B\rangle$ for scalar values x.

2. In the proof we need the Cauchy–Shcwarz inequality (notice that $|\langle A, B\rangle_F| \leq \|A\|_F\|B\|_F$ implies $\langle A, B\rangle_F \leq \|A\|_F\|B\|_F$) but is otherwise quite direct.

$$\begin{aligned} \|A+B\|_F^2 &= \langle A+B, A+B\rangle_F \\ &= \langle A, A\rangle_F + 2\langle A, B\rangle_F + \langle B, B\rangle_F \\ &\leq \|A\|_F^2 + 2\|A\|_F\|B\|_F + \|B\|_F^2 \\ &= (\|A\|_F + \|B\|_F)^2 \end{aligned}$$

Since both $\|A+B\|_F$ and $\|A\|_F + \|B\|_F$ are both non-negative numbers this implies the triangle inequality.

Obviously, as is the case with the Cauchy–Schwarz inequality, the triangle inequality also holds in general vector spaces with inner products.

3. First, we use the Cauchy–Schwarz inequality to obtain the following inequality

$$\begin{aligned} \|AB\|_F^2 &= \langle AB, AB\rangle_F = \mathrm{tr}((AB)^\top AB) \\ &= \mathrm{tr}(B^\top A^\top AB) = \mathrm{tr}(A^\top ABB^\top) \\ &= \mathrm{tr}(A^\top A(B^\top B)^\top) = \langle A^\top A, B^\top B\rangle_F \\ &\leq \|A^\top A\|_F\|B^\top B\|_F \end{aligned}$$

In order to obtain $\|AB\|_F \leq \|A\|_F \|B\|_F$ it therefore suffices to prove the inequality

$$\|A^\top A\|_F \leq \|A\|_F^2$$

To see this, we need to use a few properties of the matrix $B = A^\top A \in \mathbb{R}^{m\times m}$. Clearly, B is a symmetric matrix since $B^\top = (A^\top A)^\top = A^\top A = B$. Also, for all $\mathbf{x} \in \mathbb{R}^m$ we have

$$\langle B\mathbf{x}, \mathbf{x}\rangle = \langle A^\top A\mathbf{x}, \mathbf{x}\rangle = \langle A\mathbf{x}, A\mathbf{x}\rangle \geq 0$$

where $\langle \mathbf{x}, \mathbf{y}\rangle := \mathbf{x}^\top \mathbf{y}$ denotes the usual Euclidean inner product. In other words, B is a non-negative definite matrix. A property of non-negative definite matrices is that all their eigenvalues are non-negative (this is actually an equivalent definition of a non-negative definite matrix).

Let $\lambda_1, \ldots, \lambda_m \geq 0$ denote the eigenvalues of B. On the one hand we have

$$\|A\|_F^4 = \mathrm{tr}(A^\top A)^2 = \mathrm{tr}(B)^2 = \left(\sum_{i=1}^m \lambda_i\right)^2 = \sum_{i=1}^m \lambda_i^2 + \sum_{i\neq j} \lambda_i \lambda_j$$

since we know that the trace of a matrix equals the sum of its eigenvalues.

On the other hand we have

$$\|A^\top A\|_F^2 = \|B\|_F^2 = \mathrm{tr}(B^\top B) = \mathrm{tr}(B^2) = \sum_{i=1}^m \lambda_i^2$$

since B is a symmetric matrix and we know that if λ is an eigenvalue for B then λ^2 is an eigenvalue for B^2.

Combining the last two identities and considering that $\sum_{i\neq j} \lambda_i \lambda_j \geq 0$ because all the eigenvalues are non-negative, we can write

$$\begin{aligned} \|A^\top A\|_F^2 &= \sum_{i=1}^m \lambda_i^2 \\ &\leq \sum_{i=1}^m \lambda_i^2 + \sum_{i\neq j} \lambda_i \lambda_j \\ &= \|A\|_F^4 \end{aligned}$$

from which the desired inequality follows.

4. Using the properties of the Kronecker product we can verify the identity directly.

$$\begin{aligned} \|A\otimes B\|_F^2 &= \mathrm{tr}((A\otimes B)^\top (A\otimes B)) \\ &= \mathrm{tr}((A^\top \otimes B^\top)(A\otimes B)) \\ &= \mathrm{tr}(A^\top A \otimes B^\top B) \\ &= \mathrm{tr}(A^\top A)\mathrm{tr}(B^\top B) \\ &= \|A\|_F^2 \|B\|_F^2 \end{aligned}$$

Solution to problem 1.9, page 5: We simply write the orthogonality condition $\langle I, A\rangle_{\mathrm{F}} = 0$ for $A \in I^{\perp}$ in terms of the coordinates of

$$A = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$$

The equation is

$$\langle I, A\rangle_{\mathrm{F}} = \mathrm{tr}(I^{\mathsf{T}}A) = \mathrm{tr}(A) = x_1 + x_4 = 0$$

This gives us three 'free' variables x_2,x_3 and x_4, which agrees with the fact that $\mathbb{R}^{2\times 2}$ is a four-dimensional space which implies that the orthogonal complement of any non-zero element should be three-dimensional.

By choosing the appropriate values for the free variables we define three linearly independent solutions

$$A_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad A_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

It is straight-forward to verify that all the matrices I,A_1,A_2,A_3 are mutually orthogonal with respect to the Frobenius product. Obviously we also have $\|A_1\|_{\mathrm{F}} = \|A_2\|_{\mathrm{F}} = \|A_3\|_{\mathrm{F}} = 1$, so these three matrices form an ONB for $I^{\perp}$.

Solution to problem 1.10, page 5: The Eckart-Young theorem states that the problem of the best rank k approximation of a rank n matrix (with regard to the Frobenius norm) can be found using the SVD matrix decomposition.

The SVD of a matrix $A \in \mathbb{R}^{n\times m}$ is a matrix factorization $A = U\Sigma V^{\mathsf{T}}$ where

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \end{bmatrix} \in \mathbb{R}^{n\times m}$$

is a diagonal matrix containing the so-called 'singular' values $\sigma_1, \dots, \sigma_{\min(n,m)}$ along the diagonal and

$$U = \left[\mathbf{u}_1, \dots, \mathbf{u}_n\right] \in \mathbb{R}^{n\times n} \text{ and } V = \left[\mathbf{v}_1, \dots, \mathbf{v}_m\right] \in \mathbb{R}^{m\times m}$$

are orthogonal matrices. Another useful way of writing the SVD is the sum

$$A = \sum_{i=1}^{\min(n,m)} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathsf{T}}$$

Each term in this sum is a rank 1 matrix (every column in a matrix of the form $\sigma \mathbf{u}\mathbf{v}^{\mathsf{T}}$ is clearly a multiple of $\mathbf{u}$), these are the 'singular' matrices that give the 'singular value decomposition' its name. This sum also shows that the SVD is not completely unique. For instance, reordering the sum or changing the sign of two elements in a triple $\{\sigma_i, \mathbf{u}_i, \mathbf{v}_i\}$ does not affect the 'singular' matrices in the decomposition. However, for the standard form of the SVD we require the singular values are positive and ordered $\sigma_1 \geq \sigma_2 \geq \ldots \geq 0$.

Note also that the Frobenius norm of a 'singular' matrix in the decomposition equals

$$\|\sigma \mathbf{u}\mathbf{v}^{\mathsf{T}}\|_{\mathrm{F}}^2 = \mathrm{tr}((\sigma \mathbf{u}\mathbf{v}^{\mathsf{T}})^{\mathsf{T}}\sigma \mathbf{u}\mathbf{v}^{\mathsf{T}}) = \sigma^2 \mathrm{tr}(\mathbf{v}\mathbf{u}^{\mathsf{T}}\mathbf{u}\mathbf{v}^{\mathsf{T}}) = \sigma^2 \mathrm{tr}(\mathbf{v}\mathbf{v}^{\mathsf{T}}) = \sigma^2 \mathrm{tr}(\mathbf{v}^{\mathsf{T}}\mathbf{v}) = \sigma^2$$

since the vectors $\mathbf{u}$ are $\mathbf{v}$ are normed, $\mathbf{u}^\mathsf{T}\mathbf{u} = \mathbf{v}^\mathsf{T}\mathbf{v} = 1$. Hence, the Frobenius norm of any matrix A can be expressed in only in terms of its singular values as $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots}$. The Eckart-Young theorem states that in order to find the best rank k approximation to a matrix A we simply pick out the 'largest' (with regards to the Frobenius norm) k 'singular' matrices in the decomposition.

In order words, we form the diagonal matrix Σ_k which contains only the first k singular values (assuming $\sigma_1 \geq \sigma_2 \geq \ldots$) along the diagonal and compute $M = U\Sigma_k V^\mathsf{T}$. Among all rank k matrices M will then be the matrix that minimises the value $\|A - M\|_\mathrm{F}$.

A difficulty in using the Eckart-Young theorem when dealing with matrices without numerical computation software, is that computing the SVD for general matrices by hand can be quite tedious. For the examples below we don't use any general algorithm for computing the SVD because for special cases matrices the decomposition can often be found by simpler means.

(a) For a diagonal matrix D we basically already have the SVD $D = U\Sigma V^\mathsf{T}$ with $\Sigma = D$ and $U = V = I$. We can also explicitly write

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{bmatrix} = 2 \cdot \mathbf{e}_1\mathbf{e}_1^\mathsf{T} - 3 \cdot \mathbf{e}_2\mathbf{e}_2^\mathsf{T} + 1 \cdot \mathbf{e}_3\mathbf{e}_3^\mathsf{T}$$

The best rank 1 approximation to A contains the 'largest' term (by absolute value) in the sum

$$M = -3 \cdot \mathbf{e}_2\mathbf{e}_2^\mathsf{T} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Similarly, the best rank 2 approximation contains the largest 2 terms:

$$M = 2 \cdot \mathbf{e}_1\mathbf{e}^\mathsf{T} - 3 \cdot \mathbf{e}_2\mathbf{e}_2^\mathsf{T} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Of course, the SVD with $\Sigma = D$ and $U = V = I$ is not the standard SVD with positive and ordered singular values along the diagonal of Σ. But we could obtain this with permutation matrices (with an additional change of sign because of the -3) for U and V. For instance, we could also write

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} =: U\Sigma V^\mathsf{T}$$

to obtain the standard SVD and then compute

$$M = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} =: U\Sigma V^\mathsf{T} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which yields the same result.

(b) Noticing that B is a symmetric, we remember that a symmetric matrix can be diagonalised using an orthonormal basis, meaning we have a eigenvalue decomposition $A = QDQ^\mathsf{T}$ with Q being an orthogonal matrix. Comparing such an eigenvalue decomposition with the SVD, we notice that we can write $B = U\Sigma V^\mathsf{T}$ with $\Sigma = D$ and $U = V = Q$. Depending on the order of the eigenvalues in D we may need to permute the columns of U and V, or even change the signs of some of the columns in U or V if we have a negative eigenvalue, in order to obtain the standard SVD. But as seen in the previous example, this is not essential for the use of the Eckart-Young theorem.

To find the eigenvalue decomposition of B we first compute the characteristic polynomial

$$\det(B - \lambda I) = \begin{vmatrix} 1-\lambda & 3 \\ 3 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 9 = (-2-\lambda)(4-\lambda)$$

In order to obtain the best rank 1 approximation we actually only need the largest eigenvalue (by absolute value) $\lambda = 4$ along with the appropriate (normalised) eigenvector. Gaussian elimination

$$A - 4I = \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$

reduces the equations for the coordinates of the eigenvector $\mathbf{v} = [x_1, x_2]^\mathsf{T}$ to $x_1 - x_2 = 0$. A normalised solution is

$$\mathbf{q} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and the best rank 1 approximation can be expressed by

$$M = 4 \cdot \mathbf{q}\mathbf{q}^\mathsf{T} = 4 \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

(c) Since C is a diagonal matrix we can immediately write at least two different best rank 1 approximations to C.

$$M = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \text{ or } M = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

because we have two equal 'largest' singular values $\sigma_1 = \sigma_2 = 2$.

In fact, the are many more solutions, because in the case of two or more equal singular values there are also infinitely many valid SV decompositions. Indeed, for any orthogonal matrix $Q = U = V$ we can write a SVD for C by

$$C = 2I = Q2IQ^\mathsf{T}$$

For instance, if we choose Q to be a rotation matrix which can be written as

$$Q = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix}$$

for some $t \in \mathbb{R}$, we can even explicitly compute

$$M = Q\begin{bmatrix}2 & 0\\ 0 & 0\end{bmatrix}Q^{\mathsf{T}} = \begin{bmatrix}\cos(t) & -\sin(t)\\ \sin(t) & \cos(t)\end{bmatrix}\begin{bmatrix}2 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\cos(t) & \sin(t)\\ -\sin(t) & \cos(t)\end{bmatrix} = 2\begin{bmatrix}\cos^2(t) & \cos(t)\sin(t)\\ \cos(t)\sin(t) & \sin^2(t)\end{bmatrix}$$

to get an infinite set of different best rank 1 approximations to C (which also includes the two previously mentioned solutions by choosing $t = 0$ and $t = \frac{\pi}{2}$). One can also explicitly compute that we have $\|C - M\|_{\mathrm{F}} = 2$ for any choice of t in the expression for M above.

Solution to problem 1.11, page 5: Assume we have eigenvalue/eigenvectors pairs for A and B, $A\mathbf{v} = \lambda\mathbf{v}$ and $B\mathbf{u} = \mu\mathbf{u}$. By the definition of the Kronecker sum and properties of the Kronecker property we have

$$\begin{aligned}(A \oplus B)(\mathbf{v} \otimes \mathbf{u}) &= (A \otimes I_n + I_m \otimes B)(\mathbf{v} \otimes \mathbf{u})\\ &= (A \otimes I_n)(\mathbf{v} \otimes \mathbf{u}) + (I_m \otimes B)(\mathbf{v} \otimes \mathbf{u})\\ &= A\mathbf{v} \otimes I_n\mathbf{u} + I_n\mathbf{v} \otimes B\mathbf{u}\\ &= \lambda(\mathbf{v} \otimes \mathbf{u}) + \mu(\mathbf{v} \otimes \mathbf{u})\\ &= (\lambda + \mu)(\mathbf{v} \otimes \mathbf{u})\end{aligned}$$

This shows the sum of eigenvalues $\lambda + \mu$ is an eigenvalue for $A \oplus B$ with eigenvector $\mathbf{v} \otimes \mathbf{u}$ which proves the claim.

This means it is possible to compute the eigensystem of $A \oplus B$ without explicitly computing the Kronecker sum simply by computing the eigensystems of A and B separately and then computing all possible sums $\lambda + \mu$ of eigenvalues together with corresponding eigenvectors $\mathbf{v} \otimes \mathbf{u}$.

For A the characteristic polynomial is

$$\det(A - \lambda I) = \begin{vmatrix}-1-\lambda & 2\\ 0 & 3-\lambda\end{vmatrix} = (-1-\lambda)(3-\lambda)$$

An eigenvector for $\lambda_1 = -1$ is $\mathbf{v}_1 = [1, 0]^{\mathsf{T}}$ and an eigenvector for $\lambda_2 = 3$ is $\mathbf{v}_2 = [1, 2]^{\mathsf{T}}$.

For B we have

$$\det(B - \mu I) = \begin{vmatrix}1-\mu & 0\\ 2 & 2-\mu\end{vmatrix} = (1-\mu)(2-\mu)$$

so the eigenvalues are $\mu_1 = 1$ and $\mu_2 = 2$ with eigenvectors $\mathbf{u}_1 = [-1, 2]^{\mathsf{T}}$ and $\mathbf{u}_2 = [0, 1]^{\mathsf{T}}$.

We can organize the pairs $\lambda_i + \mu_j, \mathbf{v}_i \otimes \mathbf{u}_j$ into a table.

$\mu_j, \mathbf{u}_j \backslash \lambda_i, \mathbf{v}_i$	$-1, \begin{bmatrix}1\\0\end{bmatrix}$	$3, \begin{bmatrix}1\\2\end{bmatrix}$
$1, \begin{bmatrix}-1\\2\end{bmatrix}$	$0, \begin{bmatrix}-1\\2\\0\\0\end{bmatrix}$	$4, \begin{bmatrix}-1\\2\\-2\\4\end{bmatrix}$
$2, \begin{bmatrix}0\\1\end{bmatrix}$	$1, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$	$5, \begin{bmatrix}0\\1\\0\\2\end{bmatrix}$

Solution to problem 1.12, page 5:

(a) Since A is a symmetric matrix, we know it can be diagonalised with an orthogonal matrix U. The characteristic polynomial is

$$\det(A-\lambda I) = \begin{vmatrix} 2-\lambda & 2 \\ 2 & -1-\lambda \end{vmatrix} = (2-\lambda)(-1-\lambda)-4 = \lambda^2-\lambda-6 = (\lambda-3)(\lambda+2)$$

For $\lambda_1 = 3$ a normalised eigenvector is $\mathbf{u}_1 = \frac{1}{\sqrt{5}}[2,1]^\mathsf{T}$ and for $\lambda_2 = -2$ we have $\mathbf{u}_2 = \frac{1}{\sqrt{5}}[-1,2]^\mathsf{T}$. The matrices D and U are therefore

$$D = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix} \text{ and } U = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$$

Note that since A is a symmetric matrix this diagonalisation also gives us the SVD which we could also write as

$$A = 3 \cdot \mathbf{u}_1\mathbf{u}_1^\mathsf{T} - 2 \cdot \mathbf{u}_2\mathbf{u}_2^\mathsf{T}$$

(b) We can directly verify that if we have $U^\mathsf{T}U = I$ then we also have

$$(U\otimes U)^\mathsf{T}(U\otimes U) = (U^\mathsf{T}\otimes U^\mathsf{T})(U\otimes U) = U^\mathsf{T}U\otimes U^\mathsf{T}U = I\otimes I$$

which shows this is an orthogonal matrix. Similarly, we can see that if we have a diagonalisation $A = UDU^\mathsf{T}$ then $(U\otimes U)(D\otimes D)(U\otimes U)^\mathsf{T}$ is a diagonalisation for $A\otimes A$ since

$$(U\otimes U)(D\otimes D)(U\otimes U)^\mathsf{T} = (U\otimes U)(D\otimes D)(U^\mathsf{T}\otimes U^\mathsf{T}) = UDU^\mathsf{T}\otimes UDU^\mathsf{T} = A\otimes A$$

(c) For the best rank 1 approximation of $A\otimes A$ we need the largest singular value and corresponding singular vectors. Since A is symmetric, so is $A\otimes A$, and its SVD is the same as eigenvalue decomposition which was given in (b). We only need the largest eigenvalue of $A\otimes A$. We know that for matrices A and B the eigenvalues of $A\otimes B$ are the products $\lambda_i\mu_j$ (where λ_i and μ_j are the eigenvalues of A and B) and the eigenvectors are $\mathbf{v}_i\otimes\mathbf{u}_j$ (where $\mathbf{v}_i$ and $\mathbf{u}_j$ are the eigenvectors of A and B). The largest eigenvalue of $A\otimes A$ is therefore $\lambda_1\cdot\lambda_i = 9$ and the corresponding eigenvector is

$$\mathbf{q}_1 = \mathbf{u}_1\otimes\mathbf{u}_1 = \frac{1}{5}\begin{bmatrix} 4 \\ 2 \\ 2 \\ 1 \end{bmatrix}$$

The best rank 1 approximation is therefore

$$M_1 = 9\cdot\mathbf{q}_1\mathbf{q}_1^\mathsf{T} = \frac{9}{25}\begin{bmatrix} 16 & 8 & 8 & 4 \\ 8 & 4 & 4 & 2 \\ 8 & 4 & 4 & 2 \\ 4 & 2 & 2 & 1 \end{bmatrix}$$

To obtain the best rank 2) approximation to $A\otimes A$ we can add the second largest (by absolute value) singular matrix in the SVD to M_1. The second largest singular value (eigenvalue) of $A \otimes A$ is $\lambda_1\lambda_2 = \lambda_2\lambda_1 = -6$. The eigenspace for the eigenvalue -6 is two-dimensional, so for the eigenvector we can take

$$\mathbf{q}_2 = \mathbf{v}_1 \otimes \mathbf{v}_2 = \frac{1}{5}\begin{bmatrix} -2 \\ 4 \\ -1 \\ 2 \end{bmatrix} \text{ or } \mathbf{q}_3 = \mathbf{v}_2 \otimes \mathbf{v}_1 = \frac{1}{5}\begin{bmatrix} -2 \\ -1 \\ 4 \\ 2 \end{bmatrix}$$

or any normalised linear combination of $\mathbf{q}_2$ and $\mathbf{q}_3$, for instance $\sin(t)\mathbf{q}_2 + \cos(t)\mathbf{q}_3$ for any choice of $t \in \mathbb{R}$. If we choose $\mathbf{q}_2$ we get

$$M_2 = M_1 - 6 \cdot \mathbf{q}_2\mathbf{q}_2^{\mathsf{T}} = M_2 - \frac{6}{25}\begin{bmatrix} 4 & -8 & 2 & -4 \\ -8 & 16 & -4 & 8 \\ 2 & -4 & 1 & -2 \\ -4 & 8 & -2 & 4 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 24 & 24 & 12 & 12 \\ 24 & -12 & 12 & -6 \\ 12 & 12 & 6 & 6 \\ 12 & -6 & 6 & -3 \end{bmatrix}$$

but as in the case of Exercise 1.10 (c) we could actually construct infinitely many rank 2 matrices M with minimal distance $\|M - A \otimes A\|_{\mathrm{F}}$.

Solution to problem 1.13, page 6: Let us check what happens if we apply the matrix $B = A \otimes A + A^2 \otimes I$ on the vector $\mathbf{v} \otimes \mathbf{u}$ where $\mathbf{v}$ and $\mathbf{u}$ are eigenvectors of A, $A\mathbf{v} = \lambda\mathbf{v}$ and $A\mathbf{u} = \mu\mathbf{u}$.

$$B(\mathbf{v}\otimes\mathbf{u}) = (A\otimes A + A^2\otimes I)(\mathbf{v}\otimes\mathbf{u}) = A\mathbf{v}\otimes A\mathbf{u} + A^2\mathbf{v}\otimes I\mathbf{u} = \lambda\mu\mathbf{v}\otimes\mathbf{u} + \lambda^2\mathbf{v}\otimes\mathbf{u} = (\lambda\mu + \lambda^2)\mathbf{v}\otimes\mathbf{u}$$

This shows that we can obtain the eigensystem for B by computing the eigenvalues λ_i and eigenvectors $\mathbf{v}_i$ of A and then computing $\lambda_i\lambda_j + \lambda_i^2$ to obtain all four eigenvalues of B together with their eigenvectors $\mathbf{v}_i \otimes \mathbf{v}_j$, $i, j = 1, 2$.

The characteristic polynomial of A is

$$\det(A - \lambda I) = \begin{vmatrix} -1-\lambda & 3 \\ 3 & -1-\lambda \end{vmatrix} = (1+\lambda)^2 - 9 = (\lambda+4)(\lambda-2)$$

Eigenvectors for $\lambda_1 = -4$ and $\lambda_2 = 2$ are $\mathbf{v}_1 = [-1, 1]^{\mathsf{T}}$ and $\mathbf{v}_2 = [1, 1]^{\mathsf{T}}$. Then the eigenvalues for B are $\mu_1 = \lambda_1(\lambda_1 + \lambda_1) = 32$, $\mu_2 = \lambda_1(\lambda_2 + \lambda_1) = 8$, $\mu_3 = \lambda_2(\lambda_1 + \lambda_2) = -4$ and $\mu_4 = \lambda_2(\lambda_2 + \lambda_2) = 8$ with eigenvectors

$$\mathbf{u}_1 = \mathbf{v}_1 \otimes \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \mathbf{v}_1 \otimes \mathbf{v}_2 = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_3 = \mathbf{v}_2 \otimes \mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{u}_4 = \mathbf{v}_2 \otimes \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Solution to problem 1.14, page 6:

(a) To prove this, we need the identity

$$\mathrm{vec}(ABC) = (C^{\mathsf{T}} \otimes A)\mathrm{vec}(B)$$

which holds for any matrices A, B and C for which the product ABC is defined. Applying the (linear) operator vec to the expression $AX + XB$ we can write

$$\mathrm{vec}(AX+XB) = \mathrm{vec}(AXI) + \mathrm{vec}(IXB) = (I\otimes A)\mathrm{vec}(X) + (B^{\mathsf{T}}\otimes I)\mathrm{vec}(X) = (B^{\mathsf{T}}\oplus A)\mathrm{vec}(X)$$

which holds by the definition of the Kronecker sum.

(b) We can answer this question quickly by considering the eigenvalues of the matrix $B^\top \oplus A$. Since both $B^\top$ and A are upper diagonal matrices, we can read their eigenvalues off their diagonals. The eigenvalues of Kronecker sum $B^\top \oplus A$ are all the possible sums of the eigenvalues of $B^\top$ and A (see Exercise 1.11), i.e. $0, 1, 4, 5$. Because 0 is an eigenvalue, the matrix $B^\top \oplus A$ is a singular matrix and the homogeneous system $(B^\top \oplus A)\text{vec}(X) = 0$ has non-trivial solutions.

(c) Here we actually compute the Kronecker sum $B^\top \oplus A$ to obtain the system matrix of our equation.

$$B^\top \oplus A = B^\top \otimes I + I \otimes A = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} + \begin{bmatrix} -1 & 2 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 2 & 0 \\ 0 & 4 & 0 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

By the way, we can confirm this matrix does indeed have the eigenvalues as claimed in (b). Gaussian elimination for the system $(B^\top \oplus A)\text{vec}(X) = \text{vec}(C)$ then yields the reduced form of the system

$$\left[\begin{array}{cccc|c} 0 & 2 & 2 & 0 & -2 \\ 0 & 4 & 0 & 2 & 2 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 5 & 5 \end{array}\right] \sim \left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 2 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right] \sim \left[\begin{array}{cccc|c} 0 & 1 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right] \sim \left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

We therefore have the equations $x_2 = 0$, $x_3 = -1$ and $x_4 = 1$ for the coordinates of the unknown matrix

$$X = \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix}$$

with x_1 being the 'free' variable of the system. The general solution for our equation can then be written as

$$X = \begin{bmatrix} t & -1 \\ 0 & 1 \end{bmatrix} \text{ (with } t \in \mathbb{R})$$

Solution to problem 1.15, page 6:

(a) A square matrix A is invertible if and only if $N(A) = \{0\}$. This is true if and only if 0 is not an eigenvalue. Since by assumption all eigenvalues of A are not negative this holds if and only if they are all strictly positive.

(b) Let λ be an eigenvalue of A and $\mathbf{v}$ a corresponding eigenvector, $A\mathbf{v} = \lambda\mathbf{v}$. Since we are assuming the inverse A^{-1} exists, we can multiply this equation from the left by A^{-1} to get

$$A\mathbf{v} = \lambda\mathbf{v} \;\Rightarrow\; \mathbf{v} = \lambda A^{-1}\mathbf{v} \;\Rightarrow\; A^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v}$$

since we know $\lambda \neq 0$ from (a). This shows λ is an (non-zero) eigenvalue of A if and only if λ^{-1} is an eigenvalue of A^{-1}. Clearly then all the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A are positive if and only all the eigenvalues $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$ of A^{-1} are positive since inverting a number does not change its sign.

(c) A PSD matrix A (i.e. a symmetric matrix with only nonnegative eigenvalues) can be diagonalised, $A = QDQ^\top$, with

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues and Q is an orthogonal matrix. Let us define the square-root of D by

$$\sqrt{D} := \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix}$$

and define the matrix S by $S = Q\sqrt{D}Q^\top$. Then we can verify

$$S^2 = Q\sqrt{D}Q^\top Q\sqrt{D}Q^\top = Q\sqrt{D}\sqrt{D}Q^\top = QDQ^\top = A$$

Since by assumption all the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A are nonnegative, all the eigenvalues $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$ of S are also nonnegative, so S is also a PSD matrix.

Solution to problem 1.16, page 7:

(a) This can be done by computing the eigenvalues and verifying that they are nonnegative, as we will do in (b).

(b) We start by computing the characteristic polynomial

$$\begin{aligned}
\det(A - \lambda I) &= \begin{vmatrix} 2-\lambda & 3 & 1 \\ 3 & 6-\lambda & 3 \\ 1 & 3 & 2-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 2-\lambda & 3 & \lambda-1 \\ 3 & 6-\lambda & 0 \\ 1 & 3 & 1-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 2-\lambda & 3 & \lambda-1 \\ 3 & 6-\lambda & 0 \\ 3-\lambda & 6 & 0 \end{vmatrix} \\
&= (\lambda-1)\begin{vmatrix} 3 & 6-\lambda \\ 3-\lambda & 6 \end{vmatrix} \\
&= (1-\lambda)(18-(3-\lambda)(6-\lambda)) \\
&= \lambda(\lambda-1)(9-\lambda)
\end{aligned}$$

To obtain the eigenspace for $\lambda_1 = 0$ we compute

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

We choose a normalised solution to the system, $\mathbf{q}_1 = \frac{1}{\sqrt{3}}[1,-1,1]^\top$, so that we will have an orthonormal basis of eigenvectors.

For $\lambda_2 = 1$ we have

$$A - I \sim \begin{bmatrix} 1 & 3 & 1 \\ 3 & 5 & 3 \\ 1 & 3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 1 \\ 0 & -4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We choose $\mathbf{q}_2 = \frac{1}{\sqrt{2}}[-1,0,1]^\top$. For $\lambda_3 = 9$ we have

$$A - 9I \sim \begin{bmatrix} -7 & 3 & 1 \\ 3 & -3 & 3 \\ 1 & 3 & -7 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 1 \\ -7 & 3 & 1 \\ 1 & 3 & -7 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 1 \\ 0 & -4 & 8 \\ 0 & 4 & -8 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

We choose $\mathbf{q}_3 = \frac{1}{\sqrt{6}}[1,2,1]^\top$.

(c) As explained in Exercise 1.15 we define $\sqrt{D} = diag(0,1,3)$ and $Q = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3]$ and compute

$$\begin{aligned} \sqrt{A} &= Q\sqrt{D}Q^\top \\ &= 1 \cdot \mathbf{q}_2\mathbf{q}_2^\top + 3 \cdot \mathbf{q}_3\mathbf{q}_3^\top \\ &= \frac{1}{2}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{aligned}$$

Solution to problem 1.17, page 7: Following the algorithm, we denote

$$a_{11} = 1, \quad \mathbf{b} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad B = \begin{bmatrix} 8 & 2 \\ 2 & 6 \end{bmatrix}$$

Then we compute

$$L_1 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 8 & 2 \\ 2 & 6 \end{bmatrix} - \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 5 \end{bmatrix}$$

This completes the first step. In the next step we repeat the process for A_1. We denote

$$a_{11} = 4, \quad \mathbf{b} = \begin{bmatrix} 4 \end{bmatrix}, \quad B = \begin{bmatrix} 5 \end{bmatrix}$$

and compute

$$L_2 = \begin{bmatrix} 2 & 0 \\ 2 & 1 \end{bmatrix}, \quad A_2 = 5 - 4 = 1$$

The last step is to compute the square-root $\sqrt{A_2} = 1$ and to compute $L_3 = 1$. To get the final result it is actually not necessary to explicitly compute the product in (1.2). Namely, it can be verified that the same result can be obtained simply by nesting the first columns of the L_1, L_2 and L_3 matrices into one matrix

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 2 & 0 \\ -1 & 2 & 1 \end{bmatrix}$$

Solution to problem 2.1, page 9: First, recall that any $n \times n$ matrix A can be uniquely expressed as the linear combination

$$A = \sum_{i,j=1}^{n} a_{ij} E_{ij}$$

where a_{ij} is the (i,j)-entry of a and E_{ij} is a matrix which contains only zeros except for the (i,j)-entry which equals 1. This shows that the set of matrices $\{E_{ij}\ ;\ i,j = 1,\dots n\}$ forms a basis for the vector space $(\mathbb{R}^{n\times n}, +, \cdot)$ which we will call the *standard basis*. Since we have n^2 basis vectors this demonstrates that $(\mathbb{R}^{n\times n}, +, \cdot)$ is n^2 dimensional.

(a) Denote this subset by U_1 and assume $A, B \in U_1$. Since

$$\alpha A + \beta B = \alpha \sum_{i,j=1}^{n} a_{ij} E_{ij} + \beta \sum_{i,j=1}^{n} b_{ij} E_{ij} = \sum_{i,j=1}^{n} (\alpha a_{ij} + \beta b_{ij}) E_{ij}$$

the $(1,2)$-entry of $\alpha A + \beta B$ is $\alpha \cdot 0 + \beta \cdot 0 = 0$, ie. $\alpha A + \beta B \in U_1$. Hence, U_1 is a vector subspace of $\mathbb{R}^{n\times n}$.
For the basis of U_1 we simply omit E_{12} from the standard basis of $\mathbb{R}^{n\times n}$. Therefore, a basis of U_1 has one element less than the standard basis of $\mathbb{R}^{n\times n}$, ie. $\dim(U_1) = n^2 - 1$.

(b) For a subset of a vector space to be a vector subspace it must contain all scalar multiples of all its elements, including 0. The simplest argument why this subset is not a subspace is therefore because it does not contain the **0** matrix. Another argument is that this subset is not closed for addition since adding two matrices with 1 for their $(1,2)$ entries produces a matrix with 2 for its $(1,2)$ entry.

(c) This subset contains the **0** matrix and is also closed for addition since adding matrices with integer entries results in matrices with integer entries. However, this subset is not closed for scalar multiplication. For instance multiplying any non-zero integer matrix with for instance π will produce a matrix that does not have all integer entries.

(d) If we add two upper-triangular matrices, we get an upper-triangular matrix. Similarly, if we multiply an upper-triangular matrix with any scalar, the result will be an upper-triangular matrix. This subset, let us denote it with U_4, is therefore closed for addition and scalar multiplication and is clearly a subspace. To find a basis for U_4 and determine its dimension, we notice that any matrix $A \in U_4$ can be written as

$$A = \sum_{i \le j} a_{ij} E_{ij}$$

where the sum goes over all the pairs $i, j = 1, \dots, n$ with $i \le j$. So a basis for U_4 is the set $\{E_{ij} : i, j = 1, \dots, n,\ i \le j\}$. This set contains $\frac{1}{2}n(n+1)$ elements which is the dimension of U_4.

(e) Adding symmetric matrices produces a symmetric matrix, multiplying a symmetric matrix with any number also results in a symmetric matrix, so the set of symmetric matrices is clearly a vector subspace.
More formally, denote the subset of symmetric matrices by U_5 and assume $A, B \in U_5$, meaning $A^\top = A$ and $B^\top = B$. We can then easily verify that any linear combination of A and B is also a symmetric matrix since

$$(\alpha A + \beta B)^\top = \alpha A^\top + \beta B^\top = \alpha A + \beta B$$

To get an explicit basis for U_5 we can notice that any matrix $A \in U_5$ can be written as

$$A = \sum_{i=1}^{n} a_{ii} E_{ii} + \sum_{i<j} a_{ij} (E_{ij} + E_{ji})$$

Here, we separated the diagonal elements in the first sum and off-diagonal in the second sum. This allows us to directly identify a basis for U_5 as the set

$$\{E_{ii} : i = 1, \dots, n\} \cup \{E_{ij} + E_{ji} : i, j = 1, \dots, n,\ i < j\}$$

The number of elements in this basis and hence the dimension of U_5 is $n + \frac{1}{2}n(n-1) = \frac{1}{2}n(n+1)$.

(f) Like for the case of symmetric matrices, it is easy to conclude that the set of antisymmetric matrices is also subspace. Denote the set of antisymmetric matrices by U_6 and assume $A, B \in U_6$, implying $A^\top = -A$ and $B^\top = -B$. Then

$$(\alpha A + \beta B)^\top = \alpha A^\top + \beta B^\top = -(\alpha A + \beta B)$$

showing that any linear combination of antisymmetric matrices is an antisymmetric matrix. To get a basis we can notice that any antisymmetric matrix $A \in U_6$ can be written as

$$A = \sum_{i<j} a_{ij} (E_{ij} - E_{ji})$$

This sum does not include any non-zero diagonal elements since an antisymmetric matrix necessarily has only zeros on the diagonal. A possible basis is therefore the set

$$\{E_{ij} - E_{ji} : i, j = 1, \dots, n,\ i < j\}$$

and the dimension of U_6 is $\frac{1}{2}n(n-1)$.

To conclude, we mention that the vector spaces bases we identified for the case of symmetric and antisymmetric matrices together form a basis for the entire vector space $\mathbb{R}^{n\times n}$, which also agrees with the dimensions since $\frac{1}{2}n(n+1)+\frac{1}{2}n(n-1)=n^2$.

(g) The set of invertible matrices is not a vector subspace since it does not contain the **0** matrix. It is also not closed for addition, since even if A (and hence also $-A$) is an invertible matrix, the sum $A+(-A)=0$ is not invertible.

(h) The set of matrices with zero determinant is closed for multiplication since if $\det(A)=0$, then we have also

$$\det(\alpha A)=\alpha^n\det(A)=0$$

However this set is not closed for addition. A simple counterexample is

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

(i) The set of nilpotent matrices is again closed for multiplication by scalars, since $N^n=0$ implies $(\alpha N)^n=\alpha^n N^n=0$. But again, this set is not closed for addition. For instance let

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

Then we have $A^2=0$ and $B^2=0$, so both are nilpotent matrices. But for their sum

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

we have $C^2=I$ (and more generally $C^k=C$ for odd k and $C^k=I$ for even k) so it is not a nilpotent matrix.

(j) It is useful to provide a more explicit description of the set U_{10} of all nilpotent upper-triangular matrices. First, let A be a general upper-triangular matrix with elements $a_1,\dots,a_n$ on the diagonal

$$A=\begin{bmatrix}a_1 & * & \dots & *\\ 0 & a_2 & \dots & *\\ \vdots & & \ddots & \vdots\\ 0 & \dots & 0 & a_n\end{bmatrix}$$

If we compute the m-th power of such a matrix, we notice we get a matrix of the form

$$A^m=\begin{bmatrix}a_1^m & * & \dots & *\\ 0 & a_2^m & \dots & *\\ \vdots & & \ddots & \vdots\\ 0 & \dots & 0 & a_n^m\end{bmatrix}$$

This means we can have a nilpotent upper-triangular matrix only if all the diagonal elements equal zero, $a_1=\dots=a_n=0$.

Conversely, if we have an upper-triangular with zeros on the diagonal

$$A = \begin{bmatrix} 0 & * & \dots & * \\ 0 & 0 & \dots & * \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & 0 \end{bmatrix}$$

we can quite directly see that we get $A^n = 0$ (with every multiplication by A we see we lose an additional line of elements above the diagonal until none are left).
In other words the set of nilpotent upper-triangular matrices is precisely the set of strictly upper-triangular matrices, which is clearly a vector subspace.
Thus any $A \in U_{10}$ can be written as

$$A = \sum_{i<j} a_{ij} E_{ij}$$

A basis is the set

$$\{E_{ij} : i, j = 1, \dots, n,\ i < j\}$$

and the dimension is $\frac{1}{2}n(n+1)$.

(k) Assume we have two matrices $A, B \in U_{11}$ from the set of matrices with zero trace, $\text{tr}(A) = \text{tr}(B) = 0$. Then by the properties of the tr operator we see that any linear combination also has zero trace:

$$\text{tr}(\alpha A + \beta B) = \alpha \text{tr}(A) + \beta \text{tr}(B) = 0$$

Thus U_{11} is a vector subspace. One way of explicitly writing a general matrix $A \in U_{11}$ is

$$\begin{aligned} A &= \sum_{i=1}^{n-1} a_{ii} E_{ii} - (a_{11} + \dots + a_{n-1,n-1}) E_{nn} + \sum_{i \neq j} a_{ij} E_{ij} \\ &= \sum_{i=1}^{n-1} a_{ii} (E_{ii} - E_{nn}) + \sum_{i \neq j} a_{ij} E_{ij} \end{aligned}$$

A possible basis is then the set

$$\{E_{ii} - E_{nn} : i = 1, \dots, n-1\} \cup \{E_{ij} : i, j = 1, \dots, n,\ i \neq j\}$$

The dimension is $n - 1 + n^2 - n = n^2 - 1$. This also agrees with the fact that we have basically one equation $\text{tr}(A) = a_{11} + \dots + a_{nn} = 0$ for the entries of the matrix A and $n^2 - 1$ 'free' variables.

Solution to problem 2.2, page 9: Description.

(a) Proof.

(b) Basis.

Solution to problem 2.3, page 10:

(a) Recall that the axioms of a vector space $(V, +, \cdot)$ are:

(VS1) $u + v = v + u$ in $(u + v) + w = u + (v + w)$,

(VS2) there exists a *zero vector* $\mathbf{0}$ and $v + \mathbf{0} = \mathbf{0} + v = v$,

(VS3) for each $v \in V$ there exists an *inverse vector* $-v$, such that $v+(-v) = (-v) + v = \mathbf{0}$,

(VS4) $1 \cdot v = v$,

(VS5) $(\alpha\beta) \cdot v = \alpha \cdot (\beta \cdot v)$,

(VS6) $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$,

(VS7) $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$,

for all $u, v, w \in V$ and $\alpha, \beta \in \mathbb{R}$.
We need to verify that conditions (VS1)–(VS7) hold for $(\mathbb{R}^+, \oplus, \odot)$. Pick arbitrary $x, y, z \in \mathbb{R}^+$ and $\alpha, \beta \in \mathbb{R}$. Here we go:

(VS1) $x \oplus y = xy = yx = y \oplus x$, $x \oplus (y \oplus z) = x(yz) = xyz = (xy)z = (x \oplus y) \oplus z$,

(VS2) for $1 \in \mathbb{R}^+$ we have $x \oplus 1 = x \cdot 1 = x$, ie. 1 acts as the zero vector in $\mathbb{R}^+$; $\mathbf{0} = 1$ (and this is not contradictory),

(VS3) let's use the (ad hoc) notation $\ominus x = 1/x$ (this is well-defined since $x \neq 0$), then $x \oplus (\ominus x) = x \cdot \frac{1}{x} = 1$, which is the zero vector in $\mathbb{R}^+$,

(VS4) $1 \odot x = x^1 = x$,

(VS5) $(\alpha\beta) \odot x = x^{\alpha\beta} = (x^\beta)^\alpha = \alpha \odot (\beta \odot x)$,

(VS6) $(\alpha + \beta) \odot x = x^{\alpha+\beta} = x^\alpha x^\beta = (\alpha \odot x) \oplus (\beta \odot x)$,

(VS7) $\alpha \odot (x \oplus y) = (xy)^\alpha = x^\alpha y^\alpha = (\alpha \odot x) \oplus (\alpha \odot y)$,

hence $(\mathbb{R}^+, \oplus, \odot)$ is a vector space over $\mathbb{R}$.

(b) Note that every $x \in \mathbb{R}^+$ can be written as $x = e^{\log x} = (\log x) \odot e$, ie. every $x \in \mathbb{R}^+$ is some scalar multiple of e (the basis of the natural logarithm). Hence $\{e\}$ is a basis for $(\mathbb{R}^+, \oplus, \odot)$ and $\dim(\mathbb{R}^+) = |\{e\}| = 1$.

Solution to problem 2.4, page 10: We need to prove that for any $X, Y \in U$ and any $\alpha, \beta \in \mathbb{R}$ we have $\alpha X + \beta Y \in U$. So we pick arbitrary $X, Y \in U$ and $\alpha, \beta \in \mathbb{R}$. Now, $X, Y \in U$ means that $XN = NX$ and $YN = NY$, and we have

$$(\alpha X + \beta Y)N = \alpha XN + \beta YN = \alpha NX + \beta NY = N(\alpha X + \beta Y),$$

hence $\alpha X + \beta Y \in U$ and U is a vector subspace of $\mathbb{R}^{2\times 2}$.

To determine a basis, we need the actual matrix N. Write

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The equation $AN = NA$ becomes

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \therefore \quad \begin{bmatrix} b & 0 \\ d & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix} \quad \therefore \quad \begin{matrix} b = 0, & 0 = 0, \\ d = a, & 0 = b. \end{matrix}$$

Hence, a matrix $A \in U$ must be of the form

$$A = \begin{bmatrix} a & 0 \\ c & a \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = aI + cN,$$

where $a, c \in \mathbb{R}$ are arbitrary. Now, I and N are linearly independent, so $\mathcal{B}_U = \{I, N\}$ is a basis for U and $\dim U = 2$.

Solution to problem 2.5, page 10:

(a) Note that the zero polynomial is $0 = 0 \cdot x + 0$, and this is not contained in U_1, since $a = 0$. That means that U_1 does not contain the zero polynomial and U_1 is not a vector subspace of $\mathbb{R}_1[x]$.

(b) Pick $p, q \in U_2$ and $\alpha, \beta \in \mathbb{R}$. We have

$$(\alpha p + \beta q)(0) = \alpha p(0) + \beta q(0) = \alpha \cdot 0 + \beta \cdot 0 = 0,$$

ie. $\alpha p + \beta q \in U_2$ and U_2 is a vector subspace of $\mathbb{R}_2[x]$.

(c) No. This subset clearly does not contain the zero polynomial.

(d) Again, pick $p, q \in U_4$ and $\alpha, \beta \in \mathbb{R}$. Then

$$(\alpha p + \beta q)''(3) = \alpha p''(3) + \beta q''(3) = \alpha \cdot 0 + \beta \cdot 0 = 0,$$

ie. $\alpha p + \beta q \in U_4$ and U_4 is a vector subspace of $\mathbb{R}_n[x]$.

Solution to problem 2.7, page 11: The proof that $\mathbb{R}[x]$ is a vector space follows the same argument as the proof that $\mathbb{R}_n[x]$ is a vector space. It is routine and left to the reader. The (infinite) set of polynomials

$$\mathcal{B} = \{1, x, x^2, x^3, \dots\}$$

is clearly contained in $\mathbb{R}[x]$. Moreover, for any $p \in \mathbb{R}[x]$ we have

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

ie. p is a linear combination of polynomials in $\mathcal{B}$. Since the elements in $\mathcal{B}$ are linearly independent (as they are polynomials of different degrees), the set $\mathcal{B}$ is a basis for $\mathbb{R}[x]$. Now, $\mathcal{B}$ has infinitely many elements, hence $\dim \mathbb{R}[x] = \infty$.

For the subspace W; every polynomial $p \in W$ has zeroes 1 and -1, so it must be divisible by $(x-1)(x+1) = x^2 - 1$. Again, there are infinitely many linearly independent polynomials in $\mathbb{R}[x]$ with this property. Namely, the set

$$\mathcal{B}_W = \{x^2 - 1, x(x^2 - 1), x^2(x-1), \dots\}$$

is a basis for W and $\dim W = \infty$.

Perhaps, we should be more precise in dealing with infinities. The 'amount' of elements in $\mathcal{B}$ is the same as the 'amount' of elements in the set of the natural numbers $\mathbb{N}$. Actually, the map $\mathbb{N} \to \mathcal{B}$, $n \mapsto x^n$ is a bijection. (Here, 0 *is* a natural number.) So, in more precise notation,

$$\dim \mathbb{R}[x] = \aleph_0 \quad \text{and} \quad \dim W = \aleph_0,$$

where $\aleph_0 = |\mathbb{N}|$ is the cardinality of the set $\mathbb{N}$.

Solution to problem 2.8, page 11: Again, the proof that $\mathbb{R}[[x]]$ is a vector space is routine and can be copied, practically verbatim, from the answer to the previous exercise. We, again, and conveniently, leave this to the reader.

As an example, the Taylor series for e^x, namely

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

is an element of $\mathbb{R}[[x]]$, which is not a polynomial, so $\mathbb{R}[[x]]$ is a superset of $\mathbb{R}[x]$. The question of a basis for $\mathbb{R}[[x]]$ is actually subtler (and harder) than it may seem at a first glance. The basis cannot be explicitly constructed as it depends on the *axiom of choice*, which only (in one of its equivalent formulations) asserts the existence of a basis. (Note that the above expression for e^x is not a *finite* linear combination of vectors.) Nonetheless, the dimension of $\mathbb{R}[[x]]$ is

$$\dim \mathbb{R}[[x]] = \mathfrak{c} = 2^{\aleph_0} = |\mathbb{R}|,$$

which is strictly larger than $\aleph_0$.

Solution to problem 2.9, page 11: Pick $f, g \in V$, ie. $f'' + f = 0$ and $g'' + g = 0$, and $\alpha, \beta \in \mathbb{R}$. Then

$$(\alpha f + \beta g)'' + (\alpha f + \beta g) = \alpha f'' + \beta g'' + \alpha f + \beta g = \alpha(f'' + f) + \beta(g'' + g) = 0,$$

ie. $\alpha f + \beta g \in V$, so V is a vector subspace of $C^\infty(0, 2\pi)$.

To determine its basis recall that the general solution to the second order linear differential equation $y'' + y = 0$ is

$$y(x) = C_1 \cos x + C_2 \sin x,$$

ie. a linear combination of functions $\cos x$ and $\sin x$. As these two functions are linearly independent (as the reader will verify), the set

$$\mathcal{B}_V = \{\cos x, \sin x\}$$

is a basis for V.

Solution to problem 2.10, page 11:

(a) To shorten the notation, we set $A = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, so that $\tau(X) = AX + XA$. Then, for any $X, Y \in \mathbb{R}^{2\times 2}$ and any $\alpha, \beta \in \mathbb{R}$, we have

$$\begin{aligned}\tau(\alpha X + \beta Y) &= A(\alpha X + \beta Y) + (\alpha X + \beta Y)A = \alpha AX + \beta AY + \alpha XA + \beta YA \\ &= \alpha(AX + XA) + \beta(AY + YA) = \alpha\tau(X) + \beta\tau(Y),\end{aligned}$$

ie. τ is a linear map.

(b) We need to evaluate τ at each of the basis vectors, and then express that evaluation as a linear combination of basis vectors. We get

$$\begin{aligned}
\tau(E_{11}) &= \tau\left(\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\right) = \begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix} = \begin{bmatrix}2 & 1\\1 & 0\end{bmatrix}\\
&= 2E_{11} + E_{12} + E_{21},\\
\tau(E_{12}) &= \tau\left(\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}\right) = \begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix} = \begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}\\
&= E_{11} + E_{12} + E_{22},\\
\tau(E_{21}) &= \tau\left(\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}\right) = \begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix} = \begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}\\
&= E_{11} + E_{21} + E_{22},\\
\tau(E_{22}) &= \tau\left(\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix} + \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix} = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\\
&= E_{12} + E_{21},
\end{aligned}$$

hence, the matrix corresponding to τ with respect to the standard basis of $\mathbb{R}^{2\times 2}$ is

$$A_\tau = \begin{bmatrix}2 & 1 & 1 & 0\\1 & 1 & 0 & 1\\1 & 0 & 1 & 1\\0 & 1 & 1 & 0\end{bmatrix}.$$

Solution to problem 2.13, page 12:

(a) For $p, q \in \mathbb{R}_3[x]$ and $\alpha, \beta \in \mathbb{R}$ we have

$$\phi(\alpha p + \beta q) = (\alpha p + \beta q)(A) = \alpha p(A) + \beta q(A) = \alpha\phi(p) + \beta\phi(q),$$

ie. ϕ is linear.
For $1, x, x^2$, and x^3 we have, with a slight abuse of notation,

$$\phi(1) = I = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}, \qquad \phi(x^2) = A^2 = \begin{bmatrix}5 & 4\\4 & 5\end{bmatrix},$$

$$\phi(x) = A^1 = \begin{bmatrix}1 & 2\\2 & 1\end{bmatrix}, \qquad \phi(x^3) = A^3 = \begin{bmatrix}13 & 14\\14 & 13\end{bmatrix},$$

so the matrix corresponding to ϕ with respect to the bases $\{1, x, x^2, x^3\}$ and $\{E_{11}, E_{12}, E_{21}, E_{22}\}$ is

$$A_\phi = \begin{bmatrix}1 & 1 & 5 & 13\\0 & 2 & 4 & 14\\0 & 2 & 4 & 14\\1 & 1 & 5 & 13\end{bmatrix}.$$

(b) By the Cayley–Hamilton theorem, $\Delta_A(A) = 0$ (the zero matrix), ie. $\Delta_A \in \ker\phi$. Let's evaluate $\Delta_A(x)$:

$$\Delta_A(x) = \det(A - xI) = \begin{vmatrix}1-x & 2\\2 & 1-x\end{vmatrix} = x^2 - 2x - 3.$$

Of course, any multiple of $\Delta_A(x)$ will also annihilate A, in fact any polynomial that annihilates A is divisible by $\Delta_A(x)$. One possible choice for a basis of $\ker\phi$ is

$$\{\Delta_A(x), x\Delta_A(x)\} = \{x^2 - 2x - 3, x(x^2 - 2x - 3)\},$$

which means that $\dim(\ker\phi) = 2$.

(c) Note that $q(x) = x\Delta_A(x)$, so A is contained in that set. Moreover, $q(B) = 0$ for the matrix

$$B = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}.$$

It can be directly checked that $q(A+B) \neq 0$, which means that the set in question is not a vector subspace of $\mathbb{R}^{2\times 2}$. (How did we guess B? Note that $q(x) = x(x+1)(x-3)$, so any 2×2 matrix with two different eigenvalues picked from $\{-1, 0, 3\}$ will be annihilated by q.)

Solution to problem 2.12, page 12:

(a) This is direct. Assume $\alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c} = \mathbf{0}$, ie.

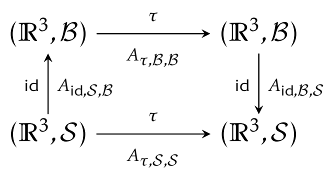
$$\alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c} = [\mathbf{a}, \mathbf{b}, \mathbf{c}]\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

A quick Gaussian elimination on the matrix $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ helps us show that $[\alpha, \beta, \gamma]^\top = [0, 0, 0]^\top$ is the unique solution to this system, hence $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are linearly independent. Since we have three linearly independent vectors in a vector space of dimension 3, they must constitute a basis of that vector space.

(b) This is also direct. Since we are already given τ on these three basis vectors, we simply read off the coefficients:

$$A_{\tau,\mathcal{B},\mathcal{B}} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(c) While we could also do this directly (and a bit tediously), we rather start with a diagram:



The nodes on this diagram represent vector spaces along with their assumed bases, while the arrows represent linear maps and the corresponding matrices with respect to the assumed bases. Note that $K := A_{\text{id},\mathcal{B},\mathcal{S}} = [\mathbf{a}, \mathbf{b}, \mathbf{c}]$, and $A_{\text{id},\mathcal{S},\mathcal{B}} = K^{-1}$. Hence, since $\tau = \text{id} \circ \tau \circ \text{id}$ (the τ on the left hand side is the bottom τ in our diagram), we have

$$A_{\tau,\mathcal{S},\mathcal{S}} = A_{\text{id},\mathcal{B},\mathcal{S}} \cdot A_{\tau,\mathcal{B},\mathcal{B}} \cdot A_{\tau,\mathcal{S},\mathcal{S}} \text{ or } A_{\tau,\mathcal{S},\mathcal{S}} = K A_{\tau,\mathcal{B},\mathcal{B}} K^{-1}.$$

After evaluating K^{-1}, and multiplying that triple product, we get

$$A_{\tau,\mathcal{S},\mathcal{S}} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

(d) The vector $[1,1,1]^{\mathsf{T}}$ is given in the standard basis, we could (with excessive use of ornaments) write

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{i} + \mathbf{j} + \mathbf{k} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{\mathcal{S}}.$$

Therefore

$$\left(\tau\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)\right)_{\mathcal{S}} = A_{\tau,\mathcal{S},\mathcal{S}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{\mathcal{S}} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}_{\mathcal{S}} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}.$$

We could also have used the matrix corresponding to τ with respect the basis $\mathcal{B}$. Note that

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{\mathcal{S}} = \frac{1}{2}(\mathbf{a} + \mathbf{b} + \mathbf{c}) = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}_{\mathcal{B}}.$$

Hence,

$$\left(\tau\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)\right)_{\mathcal{B}} = A_{\tau,\mathcal{B},\mathcal{B}} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}_{\mathcal{B}}.$$

That last column represents $\frac{3}{2}\mathbf{a} + \frac{1}{2}\mathbf{b} + \frac{1}{2}\mathbf{c}$, which is exactly $[2,2,1]^{\mathsf{T}}$ (in the standard basis).

Solution to problem 2.13, page 12:

(a) Pick arbitrary $p, q \in \mathbb{R}_3[x]$ and $\alpha, \beta \in \mathbb{R}$. Then

$$\begin{aligned} \phi(\alpha p + \beta q) &= \begin{bmatrix} (\alpha p + \beta q)(-1) \\ (\alpha p + \beta q)(0) \\ (\alpha p + \beta q)(1) \end{bmatrix} = \begin{bmatrix} \alpha p(-1) + \beta q(-1) \\ \alpha p(0) + \beta q(0) \\ \alpha p(1) + \beta q(1) \end{bmatrix} \\ &= \alpha \begin{bmatrix} p(-1) \\ p(0) \\ p(1) \end{bmatrix} + \beta \begin{bmatrix} q(-1) \\ q(0) \\ q(1) \end{bmatrix} = \alpha\phi(p) + \beta\phi(q), \end{aligned}$$

ie. ϕ is linear.

(b) For any $p \in \ker\phi$, by definition, $\phi(p) = \mathbf{0}$ must hold. In our case that means that

$$\phi(p) = \begin{bmatrix} p(-1) \\ p(0) \\ p(1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

or $p(-1) = 0$, $p(0) = 0$, $p(1) = 0$, ie. -1, 0, and 1 are zeroes of an at most degree 3 polynomial p. One choice for p is $p(x) = (x+1)x(x-1) = x^3 - x$.

Any other polynomial contained in $\ker\phi$ must be a scalar multiple of that p, since there are no higher degree polynomials in $\mathbb{R}_3[x]$. Hence, $\mathcal{B}_{\ker\phi} = \{x^3 - x\}$ is one possible basis for $\ker\phi$.

Solution to problem 2.14, page 12: Linearity of ψ essentially follows from the linearity of derivation, scalar multiplication, and addition. We have

$$\begin{aligned}\psi(\alpha p + \beta q)(x) &= (x(\alpha p + \beta q)(x+1))' - 2(\alpha p + \beta q)(x)\\ &= (\alpha x p(x+1) + \beta x q(x+1))' - 2\alpha p(x) - 2\beta q(x)\\ &= \alpha(xp(x+1))' + \beta(xq(x+1))' - 2\alpha p(x) - 2\beta q(x)\\ &= \alpha\left((xp(x+1))' - 2p(x)\right) + \beta\left((xq(x+1))' - 2q(x)\right)\\ &= \alpha\psi(p)(x) + \beta\psi(q)(x)\end{aligned}$$

for any $p, q \in \mathbb{R}_2[x]$ and any $\alpha, \beta \in \mathbb{R}$, ie. ψ is linear.

To determine the matrix corresponding to ψ we evaluate ψ on 1, x, and x^2. With slight abuse of notation, we get

$$\begin{aligned}\psi(1)(x) &= (x \cdot 1)' - 2 \cdot 1 = 1 - 2 = -1,\\ \psi(x)(x) &= (x \cdot (x+1))' - 2 \cdot x = (x^2 + x)' - 2x = 2x + 1 - 2z = 1,\\ \psi(x^2)(x) &= (x \cdot (x+1)^2)' - 2 \cdot x^2 = (x^3 + 2x^2 + x)' - 2x^2 = x^2 + 4x + 1.\end{aligned}$$

So, with respect to the basis $\{1, x, x^2\}$, the matrix corresponding to ψ is

$$A_\psi = \begin{bmatrix} -1 & 1 & 1 \\ 0 & 0 & 4 \\ 0 & 0 & 1 \end{bmatrix},$$

Finally, determination of $\ker\psi$ and $\operatorname{im}\psi$ might be easier by means of $N(A_\psi)$ and $C(A_\psi)$ in this particular case. Note that A_ψ is a rank 2 matrix, with columns 2 and 3 linearly independent. Those two columns express polynomials 1 and $1 + 4x + x^2$ with respect to the standard basis. (Namely, $\psi(x)$ and $\psi(x^2)$.) Hence,

$$\mathcal{B}_{\operatorname{im}\psi} = \{1, x^2 + 4x + 1\}$$

is a basis for $\operatorname{im}\psi$ and $\operatorname{im}\phi$ is determined. A quick Gaussian elimination on A_ψ gives

$$A_\psi \to \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

ie. columns of the form $[x_2, x_2, 0]^\mathsf{T}$ are contained in $N(A_\psi)$, or $[1, 1, 0]^\mathsf{T}$ is a choice for the sole basis vector of $N(A_\psi)$. That means that

$$\mathcal{B}_{\ker\psi} = \{x + 1\}$$

is a basis for $\ker\psi$.

Solution to problem 2.15, page 12:

(a) This is direct:

$$\begin{aligned}\phi(\alpha\mathbf{x}+\beta\mathbf{y}) &= (\alpha\mathbf{x}+\beta\mathbf{y})\mathbf{a}^\top = \alpha\mathbf{x}\,\mathbf{a}^\top+\beta\mathbf{y}\,\mathbf{a}^\top\\ &= \alpha\phi(\mathbf{x})+\beta\phi(\mathbf{y}).\end{aligned}$$

This holds for any $\mathbf{x},\mathbf{y}\in\mathbb{R}^2$ and any $\alpha,\beta\in\mathbb{R}$, ie. ϕ is linear.

(b) The standard bases of $\mathbb{R}^2$ and $\mathbb{R}^{2\times 2}$ are $\left\{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right],\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right\}$ and $\{E_{11},E_{12},E_{21},E_{22}\}$, respectively. We have

$$\begin{aligned}\phi\left(\begin{bmatrix}1\\0\end{bmatrix}\right) &= \begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix} = \begin{bmatrix}1 & 1\\0 & 0\end{bmatrix} = E_{11}+E_{12},\\ \phi\left(\begin{bmatrix}0\\1\end{bmatrix}\right) &= \begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix} = \begin{bmatrix}0 & 0\\1 & 1\end{bmatrix} = E_{21}+E_{22},\end{aligned}$$

so the matrix corresponding to ϕ is

$$A_\phi = \begin{bmatrix}1 & 0\\1 & 0\\0 & 1\\0 & 1\end{bmatrix}.$$

(c) Note that A_ϕ is of rank 2, hence $\dim(\operatorname{im}\phi) = \dim(C(A_\phi)) = 2$, and, since

$$\dim(\ker\phi)+\dim(\operatorname{im}\phi) = \dim(\mathbb{R}^2) = 2,$$

we must have $\dim(\ker\phi) = 0$.

(d) The two (linearly independent) columns of A_ϕ represent matrices

$$\begin{bmatrix}1 & 1\\0 & 0\end{bmatrix}, \text{ and } \begin{bmatrix}0 & 0\\1 & 1\end{bmatrix},$$

which, of course, constitute a basis for $\operatorname{im}\phi$.

Solution to problem 2.16, page 13:

(a) Pick $z_1,z_2\in U+V$. By the definition of $U+V$, we can write z_1 and z_2 as $z_1 = u_1+v_1$ and $z_2 = u_2+v_2$ for some $u_1,u_2\in U$ and $v_1,v_2\in V$. Therefore, for scalars $\alpha_1,\alpha_2\in\mathbb{R}$,

$$\alpha_1 z_1+\alpha_2 z_2 = \alpha_1(u_1+v_1)+\alpha_2(u_2+v_2) = (\alpha_1 u_1+\alpha_2 u_2)+(\alpha_1 v_1+\alpha_2 v_2),$$

which is an element of $U+V$, since U and V are vector subspaces of W. Hence, $U+V$ is a vector subspace of W.
The proof that $U\cap V$ is also a vector subspace is left to the reader.

(b) For $(u,v),(u',v')\in U\times V$ and $\alpha\in\mathbb{R}$ define

$$\begin{aligned}(u,v)+(u',v') &:= (u+u',v+v'),\\ \text{and }\ \alpha\cdot(u,v) &:= (\alpha u,\alpha v).\end{aligned}$$

(Note that the operations on the right side are the vector space operations in U and V, while the operations on the left side are newly defined operations.)

The routine verification that $U \times V$ is a vector space with these operations is conveniently left to the reader.
Pick bases $\{u_1, u_2, \ldots, u_m\}$ and $\{v_1, v_2, \ldots, v_n\}$ of U and V, respectively. As the reader will verify, the set

$$\{(u_1, 0), (u_2, 0), \ldots, (u_m, 0), (0, v_1), (0, v_2), \ldots, (0, v_n)\}$$

is a basis of $U \times V$. Hence, $\dim(U \times V) = \dim(U) + \dim(V)$.

(c) Let's pick $(u_1, v_1), (u_2, v_2) \in U \times V$ and $\alpha_1, \alpha_2 \in \mathbb{R}$. We have

$$\begin{aligned}\phi(\alpha_1(u_1, v_1) + \alpha_2(u_2, v_2)) &= \phi(\alpha_1 u_1 + \alpha_2 u_2, \alpha_1 v_1 + \alpha_2 v_2) \\ &= (\alpha_1 u_1 + \alpha_2 u_2) - (\alpha_1 v_1 + \alpha_2 v_2) \\ &= \alpha_1(u_1 - v_1) + \alpha_2(u_2 - v_2) \\ &= \alpha_1 \phi(u_1, v_1) + \alpha_2 \phi(u_2, v_2),\end{aligned}$$

ie. ϕ is linear.
To determine $\ker(\phi)$ we solve $\phi(u, v) = 0$ or $u - v = 0$, which means $u = v$. Note that, since $u \in U$ and $v \in V$, this also implies $u = v \in U \cap V$. Therefore

$$\ker(\phi) = \{(z, z) : z \in U \cap V\}.$$

The image of ϕ consists of vectors $\phi(u, v) = u - v$ for $u \in U$ and $v \in V$. We can rewrite this as $\phi(u, -v) = u + v$, ie. $\operatorname{im}(\phi) = U + V$.

(d) Linearity of ψ is obvious. Since $\psi(x) = (0, 0)$ implies $x = 0$, ψ is injective. Surjectivity of ψ is clear from the description of $\ker\phi$ above.

(e) We have

$$\begin{aligned}\dim(U + V) + \dim(U \cap V) = \dim(\operatorname{im}\phi) + \dim(\ker\phi) &= \dim(U \times V) \\ &= \dim(U) + \dim(V).\end{aligned}$$

Solution to problem 3.1, page 15: The Jacobian matrix is

$$J_{\mathbf{F}} = \begin{bmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{bmatrix}$$

and the Jacobi determinant is

$$\det(J_{\mathbf{F}}) = \begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix} = r\cos^2\varphi + r\sin^2\varphi = r$$

Solution to problem 3.2, page 15: The Jacobian matrix is

$$J_{\mathbf{F}} = \begin{bmatrix} \cos\varphi & -r\sin\varphi & 0 \\ \sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The Jacobian determinant is the same as for the case of polar coordinates

$$\det(J_{\mathbf{F}}) = \begin{vmatrix} \cos\varphi & -r\sin\varphi & 0 \\ \sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix} = r$$

Solution to problem 3.3, page 15: The Jacobian matrix is

$$J_{\mathbf{F}} = \begin{bmatrix} \cos\theta\cos\varphi & -r\cos\theta\sin\varphi & -r\sin\theta\cos\varphi \\ \cos\theta\sin\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta & r\cos\theta & 0 \end{bmatrix}$$

Computation of the Jacobian determinant requires a little more work than for polar and cylindrical coordinates.

$$\begin{aligned}
\det(J_{\mathbf{F}}) &= \begin{vmatrix} \cos\theta\cos\varphi & -r\cos\theta\sin\varphi & -r\sin\theta\cos\varphi \\ \cos\theta\sin\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta & 0 & r\cos\theta \end{vmatrix} \\
&= r^2 \begin{vmatrix} \cos\theta\cos\varphi & -\cos\theta\sin\varphi & -\sin\theta\cos\varphi \\ \cos\theta\sin\varphi & \cos\theta\cos\varphi & -\sin\theta\sin\varphi \\ \sin\theta & 0 & \cos\theta \end{vmatrix} \\
&= r^2\left(\sin\theta \begin{vmatrix} -\cos\theta\sin\phi & -\sin\theta\cos\varphi \\ \cos\theta\cos\phi & -\sin\theta\sin\varphi \end{vmatrix} + \cos\theta \begin{vmatrix} \cos\theta\cos\varphi & -\cos\theta\sin\phi \\ \cos\theta\sin\varphi & \cos\theta\cos\phi \end{vmatrix}\right) \\
&= r^2\left(\sin^2\theta\cos\theta \begin{vmatrix} -\sin\varphi & -\cos\phi \\ \cos\varphi & -\sin\phi \end{vmatrix} + \cos^3\theta \begin{vmatrix} \cos\varphi & -\sin\phi \\ \sin\varphi & \cos\phi \end{vmatrix}\right) \\
&= r^2\cos\theta\left(\sin^2\theta(\sin^2\varphi + \cos^2\varphi) + \cos^2\theta(\cos^2\varphi + \sin^2\varphi)\right) \\
&= r^2\cos\theta(\sin^2\theta + \cos^2\theta) \\
&= r^2\cos\theta
\end{aligned}$$

Solution to problem 3.5, page 16:

(a) With a rectangular domain of integration, this one can be directly evaluated

$$\begin{aligned}
\iint\limits_{[0,1]\times[0,1]} (5-x-y)\,dx\,dy &= \int_0^1 \left(\int_0^1 (5-x-y)\,dx\right) dy = \int_0^1 \left(5x - \frac{x^2}{2} - xy\right)\Bigg|_{x=0}^{x=1} dy \\
&= \int_0^1 \left(\frac{9}{2} - y\right) dy = \left(\frac{9y}{2} - \frac{y^2}{2}\right)\Bigg|_{y=0}^{y=1} = 4.
\end{aligned}$$

(b) From $x^2 + y^2 = 2$ we get $y = \pm\sqrt{2-x^2}$ and, since $y \geq x \geq 0$, only the solution $y = \sqrt{2-x^2}$ is relevant. We'll integrate along y first. Our integral becomes

$$\begin{aligned}
\iint_D \frac{y}{x+1}\,dx\,dy &= \int_0^1 \left(\int_x^{\sqrt{2-x^2}} \frac{y}{x+1}\,dy\right) dx = \int_0^1 \frac{y^2}{2(x+1)}\Bigg|_{y=x}^{y=\sqrt{2-x^2}} dx \\
&= \int_0^1 \frac{2-x^2-x^2}{2(x+1)}\,dx = \int_0^1 \frac{(1+x)(1-x)}{1+x}\,dx \\
&= \int_0^1 (1-x)\,dx = \frac{1}{2}.
\end{aligned}$$

(c) Let's try directly as the boundaries are given, ie. along y first (with $0 \le y \le x$) and then along x (with $0 \le x \le \pi$). We have

$$\iint_D \frac{\sin x}{x}\,dxdy = \int_0^{\pi}\left(\int_0^{x}\frac{\sin x}{x}\,dy\right)dx = \int_0^{\pi}\frac{\sin x}{x}\cdot y\Big|_{y=0}^{y=x}dx$$
$$= \int_0^{\pi}\frac{\sin x}{x}\cdot x\,dx = \int_0^{\pi}\sin x\,dx = 2.$$

(d) Since our domain of integration is whole $\mathbb{R}^2$, the boundaries for x and y are $\pm\infty$. We have

$$\iint_{\mathbb{R}^2} e^{-x^2-y^2}\,dxdy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-x^2}e^{-y^2}\,dx\right)dy = \int_{-\infty}^{\infty} e^{-y^2}\left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)dy$$
$$= \left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2}\,dy\right) = \left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)^2.$$

So, denoting $K = \iint_{\mathbb{R}^2} e^{-x^2-y^2}\,dxdy$ and $L = \int_{-\infty}^{\infty} e^{-x^2}\,dx$, we have $K = L^2$. We'll evaluate K using polar coordinates

$$x = r\cos\varphi\,,\quad y = r\sin\varphi\,,\quad \det(J_{\mathbf{F}}) = r.$$

The boundaries for r and φ are $0 \le r < \infty$ and $0 \le \varphi \le 2\pi$. We obtain

$$K = \int_0^{\infty}\left(\int_0^{2\pi} e^{-r^2\cos^2\varphi - r^2\sin^2\varphi}\cdot r\,d\varphi\right)dr = \int_0^{\infty}\left(\int_0^{2\pi} re^{-r^2}\,d\varphi\right)dr$$
$$= 2\pi\int_0^{\infty} re^{-r^2}\,dr = 2\pi\int_0^{\infty}\frac{e^{-t}}{2}\,dt = \pi,$$

where we substituted a new variable $t = r^2$ into the (single) integral in the second line.
Finally,

$$L = \int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{K} = \sqrt{\pi}.$$

Solution to problem 3.7, page 16: Let's determine the domain of integration: The projection of the intersection of the paraboloid $z = 8 - x^2 - y^2$ and the plane $z = -1$ onto the xy-plane is the curve given by

$$8 - x^2 - y^2 = -1 \quad \therefore \quad x^2 + y^2 = 9,$$

ie. a circle of radius 3 centered at the origin. The domain of integration, call it D, is a closed disk bounded by that circle. We can express the volume of the solid as a double integral in polar coordinates $x = r\cos\varphi$, $y = r\sin\varphi$, $\det(J_{\mathbf{F}}) = r$:

$$\iint_D \big(8 - x^2 - y^2 - (-1)\big)\,dxdy = \int_0^{3}\left(\int_0^{2\pi}(9 - r^2)r\,d\varphi\right)dr = 2\pi\int_0^{3}(9r - r^3)\,dr$$
$$= 2\pi\left(\frac{9r^2}{2} - \frac{r^4}{4}\right)\Big|_{r=0}^{r=3} = \frac{81\pi}{2}.$$

Solution to problem 3.8, page 16: Polar coordinates are an ideal choice for this problem. The boundaries for r and φ for that quarter of a disk D in polar coordinates are $0 \le r \le R$ and $0 \le \varphi \le \frac{\pi}{2}$. Moreover,

$$\rho(x,y) = \sqrt{x^2+y^2} = \sqrt{r^2\cos^2\varphi + r^2\sin^2\varphi} = r.$$

Hence, the integral for the mass of D is

$$\begin{aligned} m &= \iint_D \sqrt{x^2+y^2}\,dxdy = \int_0^R \left(\int_0^{\frac{\pi}{2}} r \cdot r\,d\varphi \right) dr = \frac{\pi}{2}\int_0^R r^2\,dr \\ &= \frac{\pi r^3}{6}\bigg|_{r=0}^{r=R} = \frac{\pi R^3}{6}. \end{aligned}$$

The coordinates of the center of mass are

$$\begin{aligned} x^* &= \frac{1}{m}\iint_D x\sqrt{x^2+y^2}\,dxdy = \frac{6}{\pi R^3}\int_0^R \left(\int_0^{\frac{\pi}{2}} r^2\cos\varphi \cdot r\,d\varphi \right) dr \\ &= \frac{6}{\pi R^3}\left(\int_0^R r^3\,dr \right)\left(\int_0^{\frac{\pi}{2}} \cos\varphi\,d\varphi \right) = \frac{6}{\pi R^3}\cdot\frac{R^4}{4}\cdot 1 = \frac{3R}{2\pi}, \end{aligned}$$

and

$$\begin{aligned} y^* &= \frac{1}{m}\iint_D y\sqrt{x^2+y^2}\,dxdy = \frac{6}{\pi R^3}\int_0^R \left(\int_0^{\frac{\pi}{2}} r^2\sin\varphi \cdot r\,d\varphi \right) dr \\ &= \frac{6}{\pi R^3}\left(\int_0^R r^3\,dr \right)\left(\int_0^{\frac{\pi}{2}} \sin\varphi\,d\varphi \right) = \frac{6}{\pi R^3}\cdot\frac{R^4}{4}\cdot 1 = \frac{3R}{2\pi}. \end{aligned}$$

Solution to problem 3.9, page 17: The surface $z^2 = x^2 + y^2$ is an infinite cone with (double) apex at the origin, while the surface $x^2 + y^2 + z^2 = 4$ is a sphere of radius 2 centered at the origin. Since $z \ge 0$, our domain of integration D is the 'top' conical cutout from the ball of radius 2 centered at the origin. In spherical coordinates, that domain is given by boundaries

$$0 \le r \le 2\,,\ 0 \le \varphi \le 2\pi\,,\ \frac{\pi}{4} \le \theta \le \frac{\pi}{2}.$$

Hence,

$$\begin{aligned} m &= \iiint_D \rho(x,y,z)\,dxdydz = \int_0^2 \left(\int_0^{2\pi} \left(\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} r^2\cos\theta\,d\theta \right) d\varphi \right) dr \\ &= \left(\int_0^2 r^2\,dr \right)\left(\int_0^{2\pi} d\varphi \right)\left(\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \cos\theta\,d\theta \right) = \left(\frac{r^3}{3}\bigg|_{r=0}^{r=2} \right)\cdot 2\pi\cdot\sin\theta\bigg|_{\theta=\frac{\pi}{4}}^{\theta=\frac{\pi}{2}} \\ &= \frac{8}{3}\cdot 2\pi\cdot\left(1 - \frac{\sqrt{2}}{2} \right) = \frac{8\pi(2-\sqrt{2})}{3}. \end{aligned}$$

Judging from the symmetry of our domain (and the fact that ρ is constant), we deduce that $x^* = 0$ and $y^* = 0$. (Still, the reader is invited to confirm that by evaluating corresponding integrals.) For z^* we have

$$\begin{aligned}
z^* &= \frac{1}{m}\iiint_D z\rho(x,y,z)\,dx\,dy\,dz \\
&= \frac{3}{8\pi(2-\sqrt{2})}\int_0^2\left(\int_0^{2\pi}\left(\int_{\frac{\pi}{4}}^{\frac{\pi}{2}} r\sin\theta\cdot r^2\cos\theta\,d\theta\right)d\varphi\right)dr \\
&= \frac{3}{8\pi(2-\sqrt{2})}\left(\int_0^2 r^3\,dr\right)\left(\int_0^{2\pi}d\varphi\right)\left(\int_{\frac{\pi}{4}}^{\frac{\pi}{2}}\sin\theta\cos\theta\,d\theta\right) \\
&= \frac{3}{8\pi(2-\sqrt{2})}\cdot\frac{2^4}{4}\cdot 2\pi\cdot\frac{1}{4} = \frac{3(2+\sqrt{2})}{8},
\end{aligned}$$

since

$$\int_{\frac{\pi}{4}}^{\frac{\pi}{2}}\sin\theta\cos\theta\,d\theta = \frac{1}{2}\int_{\frac{\pi}{4}}^{\frac{\pi}{2}}\sin(2\theta)\,d\theta = \frac{-\cos(2\theta)}{4}\bigg|_{\theta=\frac{\pi}{4}}^{\theta=\frac{\pi}{2}} = \frac{1+0}{4} = \frac{1}{4}.$$

Solution to problem 3.10, page 17: We can rewrite the inequality $x^2+y^2+z^2 \leq 2z$ of our domain D as

$$x^2+y^2+z^2-2z+1 \leq 1 \quad \therefore \quad x^2+y^2+(z-1)^2 \leq 1,$$

ie. our domain of integration is a ball of radius 1 centered at $(0,0,1)$. If we 'plug spherical coordinates' into the inequality $x^2+y^2+z^2 \leq 2z$, we obtain

$$r^2 \leq 2r\sin\theta,$$

and, since $r \geq 0$,

$$r \leq 2\sin\theta,$$

which gives us integration boundaries for r.

Let's start with the mass of this ball. Since $\rho(x,y,z) = \sqrt{x^2+y^2+z^2} = r$, we have

$$\begin{aligned}
m &= \iiint_D \sqrt{x^2+y^2+z^2}\,dx\,dy\,dz = \int_0^{2\pi}\left(\int_0^{\frac{\pi}{2}}\left(\int_0^{2\sin\theta} r\cdot r^2\cos\theta\,dr\right)d\theta\right)d\varphi \\
&= \left(\int_0^{2\pi}d\varphi\right)\left(\int_0^{\frac{\pi}{2}}\cos\theta\cdot\frac{r^4}{4}\bigg|_{r=0}^{r=2\sin\theta}d\theta\right) = 2\pi\int_0^{\frac{\pi}{2}}4\cos\theta\sin^4\theta\,d\theta \\
&= 8\pi\int_0^1 t^4\,dt = \frac{8\pi}{5},
\end{aligned}$$

where a new variable $t = \sin\theta$ was introduced in the last line.

Now the coordinates of the center of mass; let's start with

$$z^* = \frac{1}{m}\iiint_D z\sqrt{x^2+y^2+z^2}\,dxdydz = \frac{5}{8\pi}\int_0^{2\pi}\left(\int_0^{\frac{\pi}{2}}\left(\int_0^{2\sin\theta} r\sin\theta\cdot r\cdot r^2\cos\theta\,dr\right)d\theta\right)d\varphi$$
$$= \frac{5}{8\pi}\cdot 2\pi\int_0^{\frac{\pi}{2}}\sin\theta\cos\theta\,\frac{r^5}{5}\Big|_{r=0}^{r=2\sin\theta}d\theta = 8\int_0^{\frac{\pi}{2}}\cos\theta\sin^6\theta\,d\theta$$
$$= 8\int_0^1 t^6\,dt = \frac{8}{7}.$$

Note that $x^* = 0$ and $y^* = 0$, since $\int_0^{2\pi}\cos\varphi\,d\varphi = 0$ and $\int_0^{2\pi}\sin\varphi\,d\varphi = 0$. (The reader is invited to work out the details.)

Solution to problem 3.11, page 17: From $y^2-2 \le 2-x^2$ we get $x^2+y^2 \le 4$, ie. the projection of this solid onto the xy-plane is a disk of radius 2 centered at the origin.

The volume is

$$V = \iiint_D dxdydz = \int_0^2\left(\int_0^{2\pi}\left(\int_{r^2\sin^2\varphi-2}^{2-r^2\cos^2\varphi} dz\right)d\varphi\right)dr$$
$$= \int_0^2\left(\int_0^{2\pi}(4-r^2)d\varphi\right)dr = 2\pi\left(4r-\frac{r^3}{3}\right)\Big|_{r=0}^{r=2} = \frac{32\pi}{3}.$$

For the mass, we get

$$m = \iiint_D y^2\,dxdydz = \int_0^2\left(\int_0^{2\pi}\left(\int_{r^2\sin^2\varphi-2}^{2-r^2\cos^2\varphi} r^2\sin^2\varphi\,dz\right)d\varphi\right)dr$$
$$= \int_0^2\left(\int_0^{2\pi}(4-r^2)r^2\sin^2\varphi d\varphi\right)dr = \left(\int_0^{2\pi}\sin^2\varphi\,d\varphi\right)\left(\int_0^2(4r^2-r^4)dr\right)$$
$$= \frac{1}{2}\cdot 2\pi\cdot\left(\frac{4r^3}{3}-\frac{r^5}{5}\right)\Big|_{r=0}^{r=2} = \frac{64\pi}{15}.$$

Solution to problem 3.12, page 17:

(a) Let's start with the evaluation of the gradient of f; $\mathrm{grad}(f) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]^{\mathsf{T}}$, or the two partial derivatives:

$$\frac{\partial f}{\partial x} = 3x^2-8x+2y\ ,\ \frac{\partial f}{\partial y} = 2x-2y.$$

Stationary points of f are solutions of $\mathrm{grad}(f) = \mathbf{0}$, ie.

$$\frac{\partial f}{\partial x} = 0 \quad\therefore\quad 3x^2-8x+2y = 0,$$
$$\frac{\partial f}{\partial y} = 0 \quad\therefore\quad 2x-2y = 0.$$

We get $y = x$ from the second equation and, plugging this into the first equation, we get

$$3x^2 - 6x = 0 \quad \therefore \quad 3x(x-2) = 0,$$

ie. $x_1 = 0$, $x_2 = 2$, so the stationary points of f are $T_1(0,0)$ and $T_2(2,2)$.
We'll use the Hesse matrix of f to determine the type of these stationary points:

$$H_f(x,y) = J_{\mathrm{grad}(f)}(x,y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 6x-8 & 2 \\ 2 & -2 \end{bmatrix}.$$

In particular, at the stationary points T_1 and T_2, we have

$$H_1 := H_f(0,0) = \begin{bmatrix} -8 & 2 \\ 2 & -2 \end{bmatrix} \text{ and } H_2 := H_f(2,2) = \begin{bmatrix} 4 & 2 \\ 2 & -2 \end{bmatrix}.$$

Now, $\det(H_1) = 12$ and, since the $(1,1)$-entry of H_1 is $-8 < 0$, the matrix H_1 is negative definite by the Sylvester's criterion. Hence, T_1 *is a local maximum*. For H_2, we have $\det(H_2) = -12$, so the two (real) eigenvalues of H_2 are of opposite signs, H_2 is indefinite, and T_2 *is a saddle point* (ie. not a local extremum).

(b) Let's find stationary points first:

$$\frac{\partial g}{\partial x} = (x+1)e^x = 0,$$
$$\frac{\partial g}{\partial y} = 2(y+1)e^y = 0.$$

This system is particularly simple, and $(x_1, y_1) = (-1,-1)$ is the only solution. So $T_1(-1,-1)$ is the stationary point of g.
The Hesse matrix of g is

$$H_g(x,y) = \begin{bmatrix} (x+2)e^x & 0 \\ 0 & 2(y+2)e^y \end{bmatrix},$$

which, evaluated at the stationary point T_1, is

$$H_1 := H_g(-1,-1) = \begin{bmatrix} \frac{1}{e} & 0 \\ 0 & \frac{2}{e} \end{bmatrix}.$$

Now, this is a diagonal matrix with positive diagonal entries, ie. all eigenvalues of H_1 are positive and T_1 *is a local minimum*.

(c) Again, we start with the stationary points:

$$\frac{\partial h}{\partial x} = -(1+e^y)\sin x = 0,$$
$$\frac{\partial h}{\partial y} = e^y(\cos x - y - 1) = 0.$$

First equation forces $\sin x = 0$, ie. $x_k = k\pi$ for $k \in \mathbb{Z}$. Plugging this into the second equation we get $\cos(x_k) - y - 1 = (-1)^k - y - 1 = 0$ or

$y_k = (-1)^k - 1$. (That is, $y_k = 0$ for k even and $y_k = -2$ for k odd.) There are infinitely many stationary points $T_k(k\pi, (-1)^k - 1)$, one for each $k \in \mathbb{Z}$. Now, the Hesse matrix of h is

$$H_h(x,y) = \begin{bmatrix} -(1+e^y)\cos x & -e^y \sin x \\ -e^y \sin x & e^y(\cos x - y - 2) \end{bmatrix},$$

which, at stationary points T_k evaluates to

$$H_k := H_h(k\pi, 0) = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix} \text{ for } k \text{ even,}$$
$$\text{and } H_k := H_h(k\pi, -2) = \begin{bmatrix} 1+\frac{2}{e} & 0 \\ 0 & -\frac{1}{e} \end{bmatrix} \text{ for } k \text{ odd.}$$

Hence, for k even H_k is negative definite (since all its eigenvalues are negative), and T_k *is a local maximum for k even*. For k odd H_k is indefinite (since it has positive and negative eigenvalues), so T_k *is a saddle point for k odd*.

(d) Again, we start with the stationary points:

$$\frac{\partial k}{\partial x} = 3x^2 - 3yz = 0,$$
$$\frac{\partial k}{\partial y} = 3y^2 - 3xz = 0,$$
$$\frac{\partial k}{\partial y} = 6z - 3xy = 0.$$

Third equation implies $z = \frac{xy}{2}$. We plug this into first two equations to get

$$3x^2 - 3y \cdot \frac{xy}{2} = 0 \quad \therefore \quad 3x\left(x - \frac{y^2}{2}\right) = 0,$$
$$3y^2 - 3x \cdot \frac{xy}{2} = 0 \quad \therefore \quad 3y\left(y - \frac{x^2}{2}\right) = 0.$$

From the first one of these two equations we conclude that either $x = 0$ or $x = \frac{y^2}{2}$. In case $x = 0$, we get $y = 0$ from the second equation, and, since $z = \frac{xy}{2}$, $z = 0$. We have the first stationary point $T_1(0,0,0)$. In case $x = \frac{y^2}{2}$ we have

$$3y\left(y - \frac{y^4}{8}\right) = 0 \quad \therefore \quad 3y^2\left(1 - \frac{y^3}{8}\right) = 0.$$

Since we already covered the case $y = 0$, the only remaining option is $y = 2$, and therefore $x = 2$ and $x = 2$. The second stationary point of k is $T_2(2,2,2)$.

Now, the Hesse matrix is

$$H_k(x,y,z) = \begin{bmatrix} 6x & -3z & -3y \\ -3z & 6y & -3x \\ -3y & -3x & 6 \end{bmatrix}.$$

Evaluating at stationary points $T_1(0,0,0)$ and $T_2(2,2,2)$ we have

$$H_1 := H_k(0,0,0) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 6 \end{bmatrix} \quad \text{and} \quad H_2 := H_k(2,2,2) = \begin{bmatrix} 12 & -6 & -6 \\ -6 & 12 & -6 \\ -6 & -6 & 6 \end{bmatrix}.$$

The matrix H_1 is semidefinite (but not definite), hence *the type of stationary point T_1 cannot be determined* from second derivatives alone. (And we'll leave T_1 as is—of undetermined type.) We'll use Sylvester's criterion to determine the definiteness (or lack of it) for H_2. We have

$$12 > 0\,,\ \begin{vmatrix} 12 & -6 \\ -6 & 12 \end{vmatrix} = 108 > 0\,,\ \det(H_2) = -648 < 0,$$

ie. H_2 is indefinite and T_2 *is a saddle point*.

(e) Note that the function r is not sensitive to a permutation of variables, so all three partial derivatives can be obtained from $\frac{\partial r}{\partial x}$ with a suitable permutation of variables.
Let's start with the stationary points:

$$\begin{aligned} \frac{\partial r}{\partial x} &= 2x - 2yz = 0, \\ \frac{\partial r}{\partial y} &= 2y - 2xz = 0, \\ \frac{\partial r}{\partial x} &= 2z - 2xy = 0. \end{aligned}$$

From the third equation, we have $z = xy$, therefore the first two equations become

$$\begin{aligned} x - y \cdot xy &= x(1 - y^2) = 0, \\ y - x \cdot xy &= y(1 - x^2) = 0. \end{aligned}$$

From the first equation we have that either $x = 0$ or $1 - y^2 = 0$. In case $x = 0$, we have $z = 0$ and also $y = 0$ from the second original equation. The first stationary point is $T_1(0,0,0)$. In case $y = \pm 1$ we get $\pm 1 \cdot (1 - x^2) = 0$, so $x = \pm 1$. (N.b.: We have the two possibilities $x = \pm 1$ for each of the possibilities $y = \pm 1$.) Since $z = xy$, we get four additional stationary points

$$T_2(-1,-1,1)\,,\ T_3(-1,1,-1)\,,\ T_4(1,-1,-1)\,,\text{and}\ T_5(1,1,1).$$

The Hesse matrix is

$$H_r(x,y,z) = \begin{bmatrix} 2 & -2z & -2y \\ -2z & 2 & -2x \\ -2y & -2x & 2 \end{bmatrix}.$$

Evaluating this at stationary points $T_1, \ldots, T_5$ and denoting these matrices by $H_1, \dots, H_5$ we obtain

$$H_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, H_2 = \begin{bmatrix} 2 & -2 & 2 \\ -2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}, H_3 = \begin{bmatrix} 2 & 2 & -2 \\ 2 & 2 & 2 \\ -2 & 2 & 2 \end{bmatrix},$$
$$H_4 = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & -2 \\ 2 & -2 & 2 \end{bmatrix}, H_5 = \begin{bmatrix} 2 & -2 & -2 \\ -2 & 2 & -2 \\ -2 & -2 & 2 \end{bmatrix}.$$

Now, H_1 is clearly positive definite, so T_1 *is a local minimum*. Note that $H_2, \dots, H_5$ are indefinite by the Sylvester's criterion, ie. they all have the $(1,1)$-entry positive, while $\det(H_2) = \cdots = \det(H_5) = -32$. (In fact, the matrices $H_2, \dots, H_5$ all have eigenvalues $-2, 4, 4$.) Hence, $T_2, \dots T_5$ *are saddle points*.

(f) Stationary points:

$$\frac{\partial u}{\partial x} = 3x^2 - 3y = 0,$$
$$\frac{\partial u}{\partial y} = 3y^2 - 3x = 0.$$

So, from the first equation, $y = x^2$, hence $x^4 - x = x(x^3 - 1) = 0$ from the second equation. This gives us $x_1 = 0$ and $x_2 = 1$, hence $y_1 = 0$ and $y_2 = 1$. We have two stationary points, $T_1(0,0)$ and $T_2(1,1)$.
The Hesse matrix is

$$H_u(x,y) = \begin{bmatrix} 6x & -3 \\ -3 & 6y \end{bmatrix},$$

and, evaluating at stationary points we get

$$H_1 := H_u(0,0) = \begin{bmatrix} 0 & -3 \\ -3 & 0 \end{bmatrix} \text{ and } H_2 := H_u(1,1) = \begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix}.$$

A quick application of Sylvester's criterion reveals that H_1 is indefinite, while H_2 is positive definite. Therefore, T_1 *is a saddle point*, and T_2 *is a local minimum*.

(g) Stationary points:

$$\frac{\partial v}{\partial x} = 6xy - 6x = 0,$$
$$\frac{\partial v}{\partial y} = 3x^2 + 3y^2 - 6y = 0.$$

First equation is equivalent to $6x(y-1) = 0$ so we have either $x = 0$ or $y = 1$.

- Plugging $x = 0$ into the second equation we get $3y^2 - 6y = 0$, hence $y = 0$ or $y = 2$.
- Plugging $y = 1$ into the second equation we get $3x^2 - 3 = 0$, hence $x = -1$ or $x = 1$.

All in all, we have four stationary points: $T_1(0,0)$, $T_2(0,2)$, $T_3(-1,1)$, and $T_4(1,1)$.
The Hesse matrix of v is

$$H_v(x,y) = \begin{bmatrix} 6y-6 & 6x \\ 6x & -6 \end{bmatrix},$$

which, evaluated at $T_1,\dots,T_4$ and denoted by $H_1,\dots,H_4$ become

$$H_1 = \begin{bmatrix} -6 & 0 \\ 0 & -6 \end{bmatrix},\ H_2 = \begin{bmatrix} 6 & 0 \\ 0 & -6 \end{bmatrix},\ H_3 = \begin{bmatrix} 0 & -6 \\ -6 & -6 \end{bmatrix}, \text{and } H_4 = \begin{bmatrix} 0 & 6 \\ 6 & -6 \end{bmatrix}.$$

Now, H_1 is clearly negative definite, so T_1 *is a local maximum*. The matrices H_2, H_3, H_4 are negative definite, so T_2, T_3, T_4 *are saddle points*.

Solution to problem 3.13, page 17: Given $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ let $f(\mathbf{x}) = (\mathbf{x}^\mathsf{T}\mathbf{a})(\mathbf{x}^\mathsf{T}\mathbf{b})$.

(a) The formula for $f(\mathbf{x})$ can be rewritten as

$$f(\mathbf{x}) = (\mathbf{x}^\mathsf{T}\mathbf{a})(\mathbf{b}^\mathsf{T}\mathbf{x}) = \mathbf{x}^\mathsf{T}(\mathbf{a}\mathbf{b}^\mathsf{T})\mathbf{x}.$$

Therefore, from the formula $\frac{\partial(\mathbf{x}^\mathsf{T}A\mathbf{x})}{\partial\mathbf{x}} = \mathbf{x}^\mathsf{T}(A + A^\mathsf{T})$, we get

$$\frac{\partial f}{\partial \mathbf{x}} = \mathbf{x}^\mathsf{T}(\mathbf{a}\mathbf{b}^\mathsf{T} + \mathbf{b}\mathbf{a}^\mathsf{T}),$$

and, from the formula $\frac{\partial A\mathbf{x}}{\partial \mathbf{x}} = A$, we also get

$$\frac{\partial^2 f}{\partial \mathbf{x}^2} = \frac{\partial}{\partial \mathbf{x}}\left(\frac{\partial f}{\partial \mathbf{x}}\right)^\mathsf{T} = (\mathbf{a}\mathbf{b}^\mathsf{T} + \mathbf{b}\mathbf{a}^\mathsf{T}).$$

(b) The stationary point of f is $\mathbf{0}$, ie. the only solution of $\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) = \mathbf{0}$. Since $\mathbf{a}$ and $\mathbf{b}$ are orthogonal, ie. $\mathbf{a}^\mathsf{T}\mathbf{b} = \mathbf{b}^\mathsf{T}\mathbf{a} = 0$, we have

$$(\mathbf{a}\mathbf{b}^\mathsf{T} + \mathbf{b}\mathbf{a}^\mathsf{T})\mathbf{a} = \|\mathbf{a}\|^2\mathbf{b},$$
$$(\mathbf{a}\mathbf{b}^\mathsf{T} + \mathbf{b}\mathbf{a}^\mathsf{T})\mathbf{b} = \|\mathbf{b}\|^2\mathbf{a}.$$

Let U be the vector subspace of $\mathbb{R}^n$ spanned by $\mathbf{a}$ and $\mathbf{b}$. Restricted to U with the vector space basis $\{\mathbf{a}, \mathbf{b}\}$ the Hesse matrix of f is represented by

$$\begin{bmatrix} 0 & \|\mathbf{a}\|^2 \\ \|\mathbf{b}\|^2 & 0 \end{bmatrix},$$

which has eigenvalues $\lambda_{1,2} = \pm\|\mathbf{a}\|\,\|\mathbf{b}\|$. Hence, $\frac{\partial^2 f}{\partial \mathbf{x}^2}$ is indefinite, since it has a positive and a negative eigenvalue, so $\mathbf{0}$ *is a saddle point*.

Solution to problem 3.14, page 18: We are trying to minimize the function

$$f(\mathbf{x}) = \|\mathbf{x} - \mathbf{a}_1\|^2 + \|\mathbf{x} - \mathbf{a}_2\|^2 + \cdots + \|\mathbf{x} - \mathbf{a}_k\|^2.$$

Note that $\|\mathbf{x}-\mathbf{a}_i\|^2 = (\mathbf{x}-\mathbf{a}_i)^\mathsf{T}(\mathbf{x}-\mathbf{a}_i)$, so

$$\frac{\partial f}{\partial \mathbf{x}} = 2(\mathbf{x}-\mathbf{a}_1)^\mathsf{T} + 2(\mathbf{x}-\mathbf{a}_2)^\mathsf{T} + \cdots + 2(\mathbf{x}-\mathbf{a}_k)^\mathsf{T} = 2\left(k\mathbf{x} - \sum_{i=1}^{k} \mathbf{a}_i\right)^\mathsf{T}.$$

The stationary point of f (the solution to $\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) = \mathbf{0}$) is therefore

$$\mathbf{x} = \frac{1}{k}(\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_k).$$

Note that the Hesse matrix of f is $\frac{\partial^2 f}{\partial \mathbf{x}^2} = 2kI$ (with n-fold eigenvalue $2k$), which is positive definite, so our stationary point is in fact a (local) minimum.

Solution to problem 3.15, page 18: The problem is to find the extreme values of the function $f(x,y) = 2x^2 + y^2$ constrained to the (closed and bounded) domain given by $4(x-1)^2 + y^2 \le 16$. We split this into two subtasks.

- Find the extreme values of f on the interior of the domain, ie. subject to strict inequality $4(x-1)^2 + y^2 < 16$. Extreme values on the interior can only be attained at local extrema of f, which are contained in the interior of the domain.

- Find the extreme values of f on the boundary of the domain, ie. subject to equality $4(x-1)^2+y^2 = 16$. These will be determined using the method of Lagrange multipliers.

Let's start with the interior. Stationary points of f are solutions of the system

$$\frac{\partial f}{\partial x} = 4x = 0,$$
$$\frac{\partial f}{\partial y} = 2y = 0.$$

Clearly, $x_1 = 0, y_1 = 0$ is the only solution of this system, ie. $T_1(0,0)$ is the only stationary point of f. This point is contained in the interior of our domain (since $4(0-1)^2 + 0^2 = 4 < 16$ holds).

Now the boundary. Let's rewrite the equation $4(x-1)^2 + y^2 = 16$ as

$$\underbrace{4(x-1)^2 + y^2 - 16}_{g(x,y)} = 0,$$

ie. we have rewritten the constraint as $g(x,y) = 0$. The corresponding Lagrange function is

$$L(x,y,\lambda) = f(x,y) - \lambda g(x,y) = 2x^2 + y^2 - \lambda(4(x-1)^2 + y^2 - 16).$$

Candidates for extrema on the boundary are the stationary points of this Lagrange function. (Strictly speaking, we only need the x and y components of these stationary points.) Stationary points of L are solutions of the system

$$\frac{\partial L}{\partial x} = 4x - 8\lambda(x-1) = 0,$$
$$\frac{\partial L}{\partial y} = 2y - 2\lambda y = 0,$$
$$\frac{\partial L}{\partial \lambda} = -(4(x-1)^2 + y^2 - 16) = 0.$$

(The last equation is, of course, equivalent to our constraint.) From the second equation we get $2y(1-\lambda) = 0$, which implies $y = 0$ or $\lambda = 1$.

- The case $y_{2,3} = 0$ can be plugged directly into the third equation, and we get $(x-1)^2 = 4$ or $x_2 = -1$ and $x_3 = 3$.
- In case $\lambda = 1$ we get $-4x = -8$ from the first equation, or $x_{4,5} = 2$. Plugging this into the third equation we have $y^2 = 12$ or $y_{4,5} = \pm 2\sqrt{3}$.

Summarizing, we have the following five points, which are candidates for global extrema of f on our domain:

$T(x,x)$	$T_1(0,0)$	$T_2(-1,0)$	$T_3(3,0)$	$T_4(2,-2\sqrt{3})$	$T_5(2,2\sqrt{3})$
$f(x,y)$	0	2	18	20	20

The bottom row contains the values of f evaluated at corresponding stationary points. Clearly, the smallest value, 0, is attained at $T_1(0,0)$, while the largest value, 20, is attained at two points, $T_4(2,-2\sqrt{3})$ and $T_5(2,2\sqrt{3})$.

Solution to problem 3.16, page 18: The first octant is defined by inequalities $x \geq 0$, $y \geq 0$, and $z \geq 0$. So, along with $x+y+z=5$, we have four constraints. We'll split the task into following subtasks:

- Find candidates for extrema in the interior of the triangle T. That means extrema of g with respect to one constraint, namely $x+y+z=5$. Additionally, only candidates with $x > 0$, $y > 0$, and $z > 0$ should be considered.
- Find candidates for extrema on the edges of the triangle T. That means extrema of g with respect to two constraints, $x+y+z=5$ and one of the planes $x = 0$, $y = 0$, or $z = 0$.

Let's start with the interior of T. The Lagrange function is

$$L(x,y,z,\lambda) = xy^2z^2 - \lambda(x+y+z-5).$$

It stationary points are solutions of

$$\frac{\partial L}{\partial x} = y^2z^2 - \lambda = 0,$$
$$\frac{\partial L}{\partial y} = 2xyz^2 - \lambda = 0,$$
$$\frac{\partial L}{\partial z} = 2xy^2z - \lambda = 0,$$
$$\frac{\partial L}{\partial \lambda} = -(x+y+z-5) = 0.$$

From the first three equations we have $y^2z^2 = 2xyz^2 = 2xy^2z$. Since we only need to consider candidates with $x > 0$, $y > 0$, and $z > 0$, we can safely ignore the solutions with any of the x, y, or z equal to 0. Hence:

$$y^2z^2 = 2xyz^2 \quad \therefore \quad y = 2x \text{ and}$$
$$2xyz^2 = 2xy^2z \quad \therefore \quad z = y \quad \therefore \quad z = 2x.$$

Plugging this into the equation of the constraint, we get $5x - 5 = 0$ or $x = 1$. So, $T_1(1,2,2)$ is our first candidate.

We could define Lagrange functions with two Lagrange multipliers for each of the edges of the triagle, but, due to the simplicity of the additional constraints, we don't have to. Simply plugging $x = 0$, $y = 0$, or $z = 0$ into the above L will simplify our task. A lot!

- If $x = 0$, then $L(0,y,z,\lambda) = -\lambda(y+z-5)$. Also $g(0,y,z) = 0$.
- If $y = 0$, then $L(x,0,z,\lambda) = -\lambda(x+z-5)$. Also $g(x,0,z) = 0$.
- If $z = 0$, then $L(x,y,0,\lambda) = -\lambda(x+y-5)$. Also $g(x,y,0) = 0$.

That means that g is constant (and 0) along all edges of the triangle T. At $T_1(1,2,2)$, however, the value of g is

$$g(1,2,2) = 16.$$

So, *the largest value of g on T is* 16.

Solution to problem 3.17, page 18: The equation

$$x^2 - xy + y^2 = 3$$

represents the constraint in this task. The expression we're trying to maximize is $\|\mathbf{x}\| = \sqrt{x^2+y^2}$. Dealing with square roots can be cumbersome, and we'll see that it's preferable to maximize

$$f(x,y) = x^2 + y^2$$

instead. (N.b.: If v is a multivariate function, then every stationary point of v is also a stationary point of v^2. Why?)

So we rewrite the constraint

$$x^2 - xy + y^2 - 3 = 0$$

and form the corresponding Lagrange function

$$L(x,y,\lambda) = x^2 + y^2 - \lambda(x^2 - xy + y^2 - 3).$$

Its stationary points are solutions of

$$\frac{\partial L}{\partial x} = 2x - \lambda(2x - y) = 0,$$
$$\frac{\partial L}{\partial y} = 2y - \lambda(2y - x) = 0,$$
$$\frac{\partial L}{\partial \lambda} = -(x^2 - xy + y^2 - 3) = 0.$$

This system is a bit trickier. We can assume $\lambda \neq 0$ (since $\lambda = 0$ would imply $x = 0$ and $y = 0$ from first two equations) and rewrite the first two equations as

$$\frac{1}{\lambda} = \frac{2x - y}{2x} \quad \text{and} \quad \frac{1}{\lambda} = \frac{2y - x}{2y}.$$

Therefore

$$\frac{2x - y}{2x} = \frac{2y - x}{2y} \quad \therefore \quad 2xy - 2y^2 = 2xy - 2x^2 \quad \therefore \quad x^2 = y^2 \text{ or } y = \pm x.$$

Now we plug both cases of $y = \pm x$ into the constraint (or the third equation).

- In case $y = -x$ we have $3x^2 - 3 = 0$ or $x_{1,2} = \pm 1$.
- In case $y = x$ we have $x^2 - 3 = 0$ or $x_{3,4} = \pm\sqrt{3}$.

Let's summarize with a table

$T(x,x)$	$T_1(-1,1)$	$T_2(1,-1)$	$T_3(-\sqrt{3},-\sqrt{3})$	$T_4(\sqrt{3},\sqrt{3})$
$f(x,y)$	2	2	6	6

The largest value in the bottom row is 6, so *the points T_3 and T_4 are farthest away from the origin* (at a distance $\sqrt{6}$).

Solution to problem 3.18, page 18: As usual we prefer to maximize the square of distance from the origin rather then the distance itself. The Lagrange function is

$$L(x,y,\lambda) = x^2 + y^2 - \lambda\Big((x^2 + y^2)^2 - x^3 - y^3\Big)$$

The system for the stationary points of L is:

$$\begin{aligned}
\frac{\partial L}{\partial x} &= 2x - \lambda\Big(4x(x^2 + y^2) - 3x^2\Big) = 0,\\
\frac{\partial L}{\partial y} &= 2y - \lambda\Big(4y(x^2 + y^2) - 3y^2\Big) = 0,\\
\frac{\partial L}{\partial \lambda} &= -\Big((x^2 + y^2)^2 - x^3 - y^3\Big) = 0.
\end{aligned}$$

One obvious solution is $(x,y) = (0,0)$. We can discard this solution (since this is the origin itself) by crossing out both the x and y factors in the first two equations and then express the λ variable in two ways

$$\begin{aligned}
\frac{2}{\lambda} &= 4(x^2 + y^2) - 3x\\
\frac{2}{\lambda} &= 4(x^2 + y^2) - 3y
\end{aligned}$$

By identifying these two equations we get $x = y$ (all this assuming neither x nor y equals 0), and then the constraint gives us the equation

$$4x^4 = 2x^3$$

with $x = \frac{1}{2}$ as the only non-zero solution.

However, it is possible that only one of the variables equals zero and we still get a valid solution. Assume $x = 0$ and $y \neq 0$. The constraint then reduces to the equation

$$y^4 = y^3$$

which yields $y = 1$. We can then directly verify that $(x, y) = (0, 1)$ (with $\lambda = 2$) solves our system and is therefore a stationary point.

Similarly, assuming $x \neq 0$ and $y = 0$ leads to stationary point $(x, y) = (1, 0)$.

Out of the three non-trivial stationary points $(\frac{1}{2}, \frac{1}{2})$, $(0, 1)$ and $(1, 0)$ we see that the latter two attain the maximal distance from the origin.

Solution to problem 3.19, page 18:

(a) The extremal values of f can be situated either in the (strict) interior of the disk or on its edge and we consider both possibilities separately.
First, any extremal point in the interior of the disc must be a stationary point for f. The system for the stationary points is

$$\frac{\partial f}{\partial x} = y + 1 = 0,$$
$$\frac{\partial f}{\partial y} = x - 1 = 0,$$

with solution $(x, y) = (1, -1)$. We can verify that this point lies in the interior of our disk but if we compute the Hessian matrix

$$H_f(x, y) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

we notice this is a saddle point not an extremal point (the eigenvalues $\lambda_{1,2} = \pm 1$ are of mixed signature).
So the extremal points must be on the edge of the disk, on the circle $x^2 + y^2 = 2$. The relevant Lagrangian function for the problem is

$$L(x, y, \lambda) = xy - y + x - 1 - \lambda(x^2 + y^2 - 2)$$

and the system for the stationary points is

$$\frac{\partial L}{\partial x} = y + 1 - 2\lambda x = 0,$$
$$\frac{\partial L}{\partial y} = x - 1 - 2\lambda y = 0,$$
$$\frac{\partial L}{\partial \lambda} = -(x^2 + y^2 - 2) = 0$$

From the first two equations we express 2λ

$$2\lambda = \frac{y+1}{x} \text{ and } 2\lambda = \frac{x-1}{y}$$

This leads to

$$\frac{y+1}{x} = \frac{x-1}{y} \quad \therefore \quad y^2 + y = x^2 - x$$

By adding x^2 to both sides we get the identity

$$x^2 + y^2 + y = 2x^2 - x$$

and considering the constraint $x^2 + y^2 = 2$ this gives the expression

$$y = 2x^2 - x - 2$$

By plugging this expression for y back into the constraint equation we get an equation for x.

$$x^2 + (2x^2 - x - 2)^2 = 2 \quad \therefore \quad 2x^4 - 2x^3 - 3x^2 + 2x + 1 = 0$$

This is a quartic equation which in general has quite complicated solutions. Fortunately, we can guess two integer roots $x_1 = 1$ and $x_2 = -1$. This means the polynomial is divisible by the factor $x^2 - 1$ and computing the polynomial quotient gives

$$2x^4 - 2x^3 - 3x^2 + 2x + 1 : x^2 - 1 = 2x^2 - 2x - 1$$

which is the factor from which we can compute the remaining two roots

$$x_{3,4} = \frac{1 \pm \sqrt{3}}{2}$$

We summarize all four stationary points along with the function values in a table

$T(x,x)$	$T_1(-1,1)$	$T_2(1,-1)$	$T_3(\frac{1+\sqrt{3}}{2}, \frac{-1+\sqrt{3}}{2})$	$T_4(\frac{1-\sqrt{3}}{2}, \frac{-1-\sqrt{3}}{2})$
$f(x,y)$	-4	0	$\frac{1}{2}$	$\frac{1}{2}$

.

The minimum value of f is therefore attained at T_1, while the minimum value is achieved both at T_3 and T_4.

(b) Comparing to Exercise 3.19 (a), it is clear that any extremal points computed in (a) that happen to lie in the half-plane $x \geq 0$ are also extremal points for f on the half-disk. We notice that T_3 does indeed lie in the half-plane $x \geq 0$ which means it is where the maximum of f on our half-disk. The point T_1 however does not lie in the half-disk. Since the minimum is not achieved in the interior of the half dick nor the edge $x^2 + y^2 = 2$, it must be on the edge with $x = 0$.
The Lagrangian function for f with the constraint $x = 0$ is

$$L(x, y, \lambda) = xy - y + x - 1 - \lambda x$$

The system for the stationary points is

$$\frac{\partial L}{\partial x} = y + 1 - \lambda = 0,$$
$$\frac{\partial L}{\partial y} = x - 1 = 0,$$
$$\frac{\partial L}{\partial \lambda} = -x = 0$$

This is a contradictory system which means f has no stationary points on the (whole) line $x = 0$ (this is also clear directly since $f(0, y) = -y$).

The last possibility is that the minimum is attained on the edge of both the circle $x^2+y^2=2$ and line $x=0$, i.e. the 'corner' points $T_5(0,-\sqrt{2})$ and $T_6(0,\sqrt{2})$. The values of f at these points are $\sqrt{2}$ and $-\sqrt{2}$ respectively, which means the minimum value of f on our half-disk is achieved at the T_6 'corner' point of the half-disk.

Solution to problem 3.20, page 18: In this task, the equation of the ellipsoid, rewritten as

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}-1=0,$$

represents the constraint. An inscribed box with vertices on this ellipsoid has edges of length $2x$, $2y$, and $2z$.

(a) The inscribed box has volume $V(x,y,z)=2x\cdot 2y\cdot 2z=8xyz$ and this is the function we'd like to maximize with respect to the constraint above. The corresponding Lagrange function is

$$L(x,y,z,\lambda)=8xyz-\lambda\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}-1\right).$$

The system that determines the stationary points of L is:

$$\begin{aligned}\frac{\partial L}{\partial x}&=8yz-\frac{2\lambda x}{a^2}=0,\\ \frac{\partial L}{\partial y}&=8xz-\frac{2\lambda y}{b^2}=0,\\ \frac{\partial L}{\partial z}&=8xy-\frac{2\lambda z}{c^2}=0,\\ \frac{\partial L}{\partial \lambda}&=-\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}-1\right)=0.\end{aligned}$$

Multiplying the first three equations with yz, xz, and xy respectively, and then rearranging we obtain

$$\begin{aligned}4a^2y^2z^2&=\lambda xyz,\\ 4b^2x^2z^2&=\lambda xyz,\\ 4c^2x^2y^2&=\lambda xyz,\end{aligned}$$

and, ignoring cases with $x=0$, $y=0$, or $z=0$, we deduce

$$\frac{x^2}{a^2}=\frac{y^2}{b^2}=\frac{z^2}{c^2}.$$

Pluging this into the constraint we get

$$3\cdot\frac{x^2}{a^2}=1\ ,\ 3\cdot\frac{y^2}{b^2}=1\ ,\text{ and }\ 3\cdot\frac{z^2}{c^2}=1,$$

ie. $x=\pm\frac{a}{\sqrt{3}}$, $y=\pm\frac{b}{\sqrt{3}}$, and $z=\pm\frac{c}{\sqrt{3}}$. Ignoring the signs (and solutions with any of the edge lengths equal to 0), we deduce that

$$V\left(\frac{a}{\sqrt{3}},\frac{b}{\sqrt{3}},\frac{c}{\sqrt{3}}\right)=\frac{8abc}{3\sqrt{3}}$$

is the largest possible volume of the inscribed rectangular box.

(b) We just need to replace the function to maximize, in this subtask it is

$$S(x,y,z) = 2x \cdot 2y + 2x \cdot 2z + 2y \cdot 2z = 4(xy + xz + yz).$$

As we will see, we are now presented with a slightly more formidable problem. The Lagrange function becomes

$$L(x,y,z,\lambda) = 4(xy + xz + yz) - \lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1\right).$$

Its stationary points are solutions of

$$\begin{aligned}
\frac{\partial L}{\partial x} &= 4(y+z) - \frac{2\lambda x}{a^2} = 0,\\
\frac{\partial L}{\partial y} &= 4(x+z) - \frac{2\lambda y}{b^2} = 0,\\
\frac{\partial L}{\partial z} &= 4(x+y) - \frac{2\lambda z}{c^2} = 0,\\
\frac{\partial L}{\partial \lambda} &= -\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1\right) = 0.
\end{aligned}$$

Now, we multiply the first three equations with $\frac{a^2}{2}$, $\frac{b^2}{2}$, and $\frac{c^2}{2}$, respectively, to obtain

$$\begin{aligned}
2a^2(y+z) &= \lambda x,\\
2b^2(x+z) &= \lambda y,\\
2c^2(x+y) &= \lambda z.
\end{aligned}$$

This is now an eigenvalue problem. Namely, setting

$$A = \begin{bmatrix} 0 & 2a^2 & 2a^2 \\ 2b^2 & 0 & 2b^2 \\ 2c^2 & 2c^2 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

the above system becomes $A\mathbf{x} = \lambda\mathbf{x}$. The characteristic polynomial of A is

$$\det(A - \lambda I) = -\lambda^3 + 4(a^2b^2 + a^2c^2 + b^2c^2)\lambda + 16a^2b^2c^2.$$

Note that this is a depressed cubic (ie. it has no λ^2 term) and using Cardano's formula one can deduce that the only real zero is

$$\lambda_1 = \frac{4(a^2b^2 + a^2c^2 + b^2c^2)}{C} + \frac{C}{3},$$

where

$$C = \sqrt[3]{\frac{432a^2b^2c^2 + \sqrt{186624a^4b^4c^4 - 6912(a^2b^2 + a^2c^2 + b^2c^2)^3}}{2}}.$$

The unintimidated reader can continue from here.

Solution to problem 3.21, page 19: Denote the lengths of the edges of this box by a, b, and c. Since we assembled the box frame from a rod of length ℓ, we must have $4a+4b+4c=\ell$. This is our constraint.

(a) The volume of the box is $V(a,b,c)=abc$, and this is precisely the function we must maximize, subject to the constraint above. Let's rewrite the constraint as

$$4a+4b+4c-\ell=0$$

and form the Lagrange function

$$L(a,b,c,\lambda)=abc-\lambda(4a+4b+4c-\ell).$$

The stationary point of L are solutions of the system

$$\begin{aligned}\frac{\partial L}{\partial a}&=bc-4\lambda=0,\\ \frac{\partial L}{\partial b}&=ac-4\lambda=0,\\ \frac{\partial L}{\partial c}&=ab-4\lambda=0,\\ \frac{\partial L}{\partial \lambda}&=-(4a+4b+4c-\ell)=0.\end{aligned}$$

We quickly gather from the first three equations that $bc=ac=ab$ holds. While we could safely ignore the solutions with any of the a, b, or c equal to 0, let's be strict (once) and find all solutions to this system.
If $a=0$, then we must have $\lambda=0$. (Consider either the second or the third equation.) Now, from the first equation (and since $\lambda=0$) we must have either $b=0$ or $c=0$. If $b=0$, we have $c=\frac{\ell}{4}$, if $c=0$, we have $b=\frac{\ell}{4}$.
There's nothing special about starting the reasoning above with 'if $a=0$'. The conclusion has the same form: Any solution with one of the a, b, or c equal to 0, will force $\lambda=0$, two of the a, b, and c equal to 0, and the remaining one equal to $\frac{\ell}{4}$. These solutions are

$$\left(\tfrac{\ell}{4},0,0,0\right),\ \left(0,\tfrac{\ell}{4},0,0\right),\ \text{and}\ \left(0,0,\tfrac{\ell}{4},0\right).$$

They are also not the ones that interest us—the volume of the resulting (degenerate) box is 0, ie. these solutions represent the minima.
So let's assume that none of the a, b, c are equal to 0. Then, from $bc=ac=ab$, we deduce that $b=a$ and $c=b$, ie. $a=b=c$. Plugging this into the constraint, we get $a=\frac{\ell}{12}$ and

$$V\left(\frac{\ell}{12},\frac{\ell}{12},\frac{\ell}{12}\right)=\left(\frac{\ell}{12}\right)^3$$

is the largest possible volume of such a box frame. So the box frame with largest possible volume we can assemble from a rod of length ℓ is in fact a frame of a cube.

(b) The additional restriction is in fact an additional constraint, $ab=A$ must hold. We'll rewrite this as $ab-A=0$. The Lagrange function will now depend on two Lagrange multipliers, we'll denote them by λ and μ:

$$L(a,b,c,\lambda,\mu)=abc-\lambda(4a+4b+4c-\ell)-\mu(ab-A).$$

We now solve the system:

$$\begin{aligned}
\frac{\partial L}{\partial a} &= bc - 4\lambda - \mu b = 0,\\
\frac{\partial L}{\partial b} &= ac - 4\lambda - \mu a = 0,\\
\frac{\partial L}{\partial c} &= ab - 4\lambda = 0,\\
\frac{\partial L}{\partial \lambda} &= -(4a + 4b + 4c - \ell) = 0,\\
\frac{\partial L}{\partial \lambda} &= -(ab - A) = 0.
\end{aligned}$$

From the third equation we have $4\lambda = ab$, and replacing 4λ with ab in the first two equations we get

$$\begin{aligned}
bc - ab - \mu b = 0 \quad &\therefore \quad b(c - a - \mu) = 0,\\
ac - ab - \mu a = 0 \quad &\therefore \quad a(c - b - \mu) = 0.
\end{aligned}$$

We're assuming that $A > 0$, and, since $ab = A$ (from the fifth equation), $a \neq 0$ and $b \neq 0$. Hence,

$$\begin{aligned}
c - a - \mu &= 0,\\
c - b - \mu &= 0,
\end{aligned}$$

which implies $a = b$. From the second constraint we now get $a^2 = A$ or $a = \sqrt{A}$. (We ignore the negative solution, since the length cannot be negative.) To finish, we use the first constraint (the fourth equation in our 'Lagrange system') to get $8\sqrt{A} + 4c = \ell$ or $c = \frac{\ell}{4} - 2\sqrt{A}$. The maximum possible volume in this case is therefore

$$V\left(\sqrt{A}, \sqrt{A}, \frac{\ell}{4} - 2\sqrt{A}\right) = A \cdot \left(\frac{\ell}{4} - 2\sqrt{A}\right).$$

The solution is, in a sense, expected. We obtained a box with base rectangle a square with side length $\sqrt{A}$, the remainder of the rod length is then used for four vertical edges.

Solution to problem 3.22, page 19: This is similar to the previous exercise. Denote by a the side length of the base equilateral triangle and by h the prism's height. The volume of such a prism is

$$V(a,h) = \frac{a^2 h\sqrt{3}}{4},$$

and its surface area is

$$A(a,h) = \frac{a^2\sqrt{3}}{2} + 3ah.$$

Having ℓ meters of the rod available, means that $6a + 3h = \ell$. That will be our constraint.

(a) We need to maximize V with respect to the constraint. The Lagrange function is

$$L(a,h,\lambda) = \frac{a^2h\sqrt{3}}{4} - \lambda(6a+3h-\ell).$$

It stationary points are solutions of

$$\frac{\partial L}{\partial a} = \frac{ah\sqrt{3}}{2} - 6\lambda = 0,$$
$$\frac{\partial L}{\partial b} = \frac{a^2\sqrt{3}}{4} - 3\lambda = 0,$$
$$\frac{\partial L}{\partial \lambda} = -(6a+3h-\ell) = 0.$$

Let's multiply the first equation with 2 and the second one with 4 to get

$$ah\sqrt{3} - 12\lambda = 0,$$
$$a^2\sqrt{3} - 12\lambda = 0.$$

From this we get

$$ah = a^2 \quad \therefore \quad ah - a^2 = 0 \quad \therefore \quad a(h-a) = 0.$$

Ignoring the solutions with $a = 0$, we get $h = a$. (The reader is invited to consider the solutions we just ignored. What do they represent?) Hence, from the constraint, we get $9a - \ell = 0$ or $a = h = \frac{\ell}{9}$. That means that we need to cut up the rod into 9 pieces of equal lenght, and the resulting maximal volume is

$$V\left(\frac{\ell}{9}, \frac{\ell}{9}\right) = \frac{\ell^3\sqrt{3}}{4\cdot 9^3}.$$

(b) We now need to maximize A with respect to our constraint. The Lagrange function is

$$L(a,h,\lambda) = \frac{a^2\sqrt{3}}{2} + 3ah - \lambda(6a+3h-\ell).$$

It stationary points are now solutions of

$$\frac{\partial L}{\partial a} = a\sqrt{3} + 3h - 6\lambda = 0,$$
$$\frac{\partial L}{\partial b} = 3a - 3\lambda = 0,$$
$$\frac{\partial L}{\partial \lambda} = -(6a+3h-\ell) = 0.$$

Note that $\lambda = a$ from the second equation, and, plugging this into the first equation, we get

$$a\sqrt{3} + 3h - 6a = 0 \quad \therefore \quad 3h = (6-\sqrt{3})a.$$

Substituting that instead of $3h$ into the third equation, we have

$$6a + (6-\sqrt{3})a - \ell = 0 \quad \therefore \quad a = \frac{\ell}{12-\sqrt{3}} = \frac{12+\sqrt{3}}{141}\ell \doteq 0.09739\ell,$$

and

$$3h = \frac{(6-\sqrt{3})a}{12-\sqrt{3}} \quad \therefore \quad h = \frac{(6-\sqrt{3})a}{3(12-\sqrt{3})} = \frac{23-2\sqrt{3}}{141}\ell \doteq 0.13855\ell.$$

The reader is invited to evaluate the resulting maximal attainable area.

Solution to problem 3.23, page 19:

(a) We'd like to find extreme values of the function $f(\mathbf{x}) = \mathbf{a}^\mathsf{T}\mathbf{x}$ subject to the constraint $\|\mathbf{x}\| = d$. We first rewrite the constraint as

$$\|\mathbf{x}\|^2 = d^2 \quad \therefore \quad \|\mathbf{x}\|^2 - d^2 = 0 \quad \therefore \quad \underbrace{\mathbf{x}^\mathsf{T}\mathbf{x} - d^2}_{g(\mathbf{x})} = 0$$

and set the Lagrange function as

$$L(\mathbf{x},\lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x}) = \mathbf{a}^\mathsf{T}\mathbf{x} - \lambda(\mathbf{x}^\mathsf{T}\mathbf{x} - d^2).$$

The stationary points of L are again the solutions of the system

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{a}^\mathsf{T} - 2\lambda\mathbf{x}^\mathsf{T} = \mathbf{0},$$
$$\frac{\partial L}{\partial \lambda} = -(\mathbf{x}^\mathsf{T}\mathbf{x} - d^2) = 0.$$

It follows from the first equation that $\mathbf{a} = 2\lambda\mathbf{x}$, ie. $\mathbf{x}$ and $\mathbf{a}$ are parallel. Let's write this as $\mathbf{x} = \alpha\mathbf{a}$ and plug it into the second equation:

$$\alpha^2\mathbf{a}^\mathsf{T}\mathbf{a} - d^2 = 0 \quad \therefore \quad \alpha^2 = \frac{d^2}{\|\mathbf{a}\|^2} \quad \therefore \quad \alpha = \pm\frac{d}{\|\mathbf{a}\|}.$$

Hence, vectors $\mathbf{x}$, at which extreme values of f are attained, are

$$\mathbf{x} = \alpha\mathbf{a} = \pm\frac{d}{\|\mathbf{a}\|}\mathbf{a},$$

and these extreme values are

$$f\Big(\pm\frac{d}{\|\mathbf{a}\|}\mathbf{a}\Big) = \mathbf{a}^\mathsf{T}\Big(\pm\frac{d}{\|\mathbf{a}\|}\mathbf{a}\Big) = \pm d\|\mathbf{a}\|.$$

(b) Extreme value of the expression $\mathbf{a}^\mathsf{T}\mathbf{x}$ (the dot product of $\mathbf{a}$ and $\mathbf{x}$) on the sphere with equation $\|\mathbf{x}\| = d$ will be attained precisely when $\mathbf{x}$ is parallel to $\mathbf{a}$.

Solution to problem 3.24, page 19:

(a) Let's rewrite the constraint as $\|\mathbf{x}\|^2 = d^2$ or $\mathbf{x}^\mathsf{T}\mathbf{x} - d^2 = 0$. The Lagrange function corresponding to our problem is

$$L(\mathbf{x},\lambda) = \mathbf{x}^\mathsf{T}A\mathbf{x} - \lambda(\mathbf{x}^\mathsf{T}\mathbf{x} - d^2).$$

Its stationary points are, as usual, solutions to the system

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x}^\top (A + A^\top) - 2\lambda \mathbf{x}^\top = \mathbf{0},$$
$$\frac{\partial L}{\partial \lambda} = -(\mathbf{x}^\top \mathbf{x} - d^2) = 0.$$

Note that the first (system of) equation(s) can be rewritten as

$$\frac{A + A^\top}{2}\mathbf{x} = \lambda \mathbf{x},$$

ie. $\mathbf{x}$ is an eigenvector of $\frac{A+A^\top}{2}$ (and λ is the corresponding eigenvalue). (Since $\frac{A+A^\top}{2}$ is symmetric, its eigenvalues and its eigenvectors are real, ie. $\lambda \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$.) Another thing to notice is

$$\mathbf{x}^\top \Big(\frac{A + A^\top}{2}\Big)\mathbf{x} = \frac{1}{2}\left(\mathbf{x}^\top A\mathbf{x} + \mathbf{x}^\top A^\top \mathbf{x}\right) = \frac{1}{2} \cdot 2\mathbf{x}^\top A\mathbf{x} = \mathbf{x}^\top A\mathbf{x} = f(\mathbf{x}).$$

So, for an eigenvector $\mathbf{x}$ of $\frac{A+A^\top}{2}$ with $\|\mathbf{x}\| = d$, we have

$$f(\mathbf{x}) = \mathbf{x}^\top \Big(\frac{A + A^\top}{2}\Big)\mathbf{x} = \mathbf{x}^\top \lambda \mathbf{x} = \lambda d^2.$$

Finally, the extreme values of f subject to $\|\mathbf{x}\| = d$ can be identified as $\lambda_{\max} d^2$ and $\lambda_{\min} d^2$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and the smallest eigenvalues of $\frac{A+A^\top}{2}$, respectively.

(b) Now the constraint is $\mathbf{x}^\top A\mathbf{x} = d^2$, which we rewrite as $\mathbf{x}^\top A\mathbf{x} - d^2 = 0$. For the Lagrange function we have

$$L(\mathbf{x}, \mu) = \mathbf{x}^\top \mathbf{x} - \mu(\mathbf{x}^\top A\mathbf{x} - d^2).$$

(The decision to denote the Lagrange multiplier by μ will become clear later.) Its stationary points are solutions of

$$\frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{x}^\top - 2\mu \mathbf{x}^\top A = \mathbf{0},$$
$$\frac{\partial L}{\partial \mu} = -(\mathbf{x}^\top A\mathbf{x} - d^2) = 0.$$

(We used the fact that A is symmetric when evaluating the first derivative.) Rewriting the first equation as

$$A\mathbf{x} = \frac{1}{\mu}\mathbf{x},$$

we see that $\frac{1}{\mu}$ must be an eigenvalue of A, with $\mathbf{x}$ the corresponding eigenvector. (Note that $\frac{1}{\mu}$ makes sense as an eigenvalue of A, since A is a definite matrix.) From the second equation (the constraint), we now obtain

$$\mathbf{x}^\top A\mathbf{x} = d^2 \quad \therefore \quad \mathbf{x}^\top \left(\frac{1}{\mu}\mathbf{x}\right) = d^2 \quad \therefore \quad \|\mathbf{x}\|^2 = \mu d^2 \ \text{ or } \ f(\mathbf{x}) = \mu d^2.$$

Hence, denoting the smallest and the largest eigenvalues of A by $\lambda_{\min}$ and $\lambda_{\max}$, respectively, we have that

$$\frac{d^2}{\lambda_{\min}} \text{ is the largest value attained by } f\text{, and}$$
$$\frac{d^2}{\lambda_{\max}} \text{ is the smallest value attained by } f.$$

(We used that $\lambda = \frac{1}{\mu}$ for an eigenvalue λ of A, and also the fact that the eigenvalues of A are positive.)

Solution to problem 3.25, page 19:

(a) The inequality $\|\mathbf{x}-\mathbf{p}\| \le d$ determines a closed ball of radius d centered at $\mathbf{p}$. We split the solution into two subtasks: extrema in the interior (determined by $\|\mathbf{x}-\mathbf{p}\| < d$) and extrema on the boundary (determined by $\|\mathbf{x}-\mathbf{p}\| = d$).

- The interior: This is easy, the only stationary point of f is $\mathbf{x} = \mathbf{0}$ and this is a candidate if and only of $\|\mathbf{p}\| < d$, ie. $\mathbf{0}$ is actually contained in the interior.
- The boundary: Write the constraint as

$$\|\mathbf{x}-\mathbf{p}\| = d \quad \therefore \quad \|\mathbf{x}-\mathbf{p}\|^2 - d^2 = 0 \quad \therefore \quad (\mathbf{x}-\mathbf{p})^\mathsf{T}(\mathbf{x}-\mathbf{p}) - d^2 = 0,$$

and let's set up the corresponding Lagrange function

$$L(\mathbf{x},\lambda) = \mathbf{x}^\mathsf{T}\mathbf{x} - \lambda\left((\mathbf{x}-\mathbf{p})^\mathsf{T}(\mathbf{x}-\mathbf{p}) - d^2\right).$$

Now,

$$\frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{x}^\mathsf{T} - \lambda(2\mathbf{x}^\mathsf{T} - 2\mathbf{p}^\mathsf{T}) = \mathbf{0}^\mathsf{T},$$
$$\frac{\partial L}{\partial \lambda} = -\left(\|\mathbf{x}-\mathbf{p}\|^2 - d^2\right) = 0.$$

The first equation implies that $(2-2\lambda)\mathbf{x} = -2\lambda\mathbf{p}$, ie. $\mathbf{x}$ and $\mathbf{p}$ must be parallel. Let's write this as $\mathbf{x} = \alpha\mathbf{p}$ for some $\alpha \in \mathbb{R}$ and plug this into the second equation

$$\|\alpha\mathbf{p}-\mathbf{p}\| = d \quad \therefore \quad \|(\alpha-1)\mathbf{p}\| = d \quad \therefore \quad |\alpha-1| = \frac{d}{\|\mathbf{p}\|} \quad \therefore \quad \alpha = 1 \pm \frac{d}{\|\mathbf{p}\|}.$$

Hence,

$$\mathbf{x} = \left(1 \pm \frac{d}{\|\mathbf{p}\|}\right)\mathbf{p},$$

which should be the expected solution.

Finally, if $\|\mathbf{p}\| < d$, then the minimal value of f is attained in the interior at $\mathbf{0}$, and is $f(\mathbf{0}) = 0$. In case $\|\mathbf{p}\| \ge d$, the minimal value of f is attained on the boaundary at $\left(1 - \frac{d}{\|\mathbf{p}\|}\right)\mathbf{p}$, and is equal to $\|\mathbf{p}\|^2 + d^2 - 2d\|\mathbf{p}\|$. (As a sanity check, notice that in the boundary case $\|\mathbf{p}\| = d$ that last expression is 0, as it should be.)

(b) The equation $A\mathbf{x} = \mathbf{b}$ determines an affine subspace of $\mathbb{R}^n$. The appropriate Lagrange function now is

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x}^\mathsf{T}\mathbf{x} - \boldsymbol{\lambda}^\mathsf{T}(A\mathbf{x} - \mathbf{b}).$$

(Note that $\boldsymbol{\lambda}$ is now a column vector!) Its stationary points are solutions of

$$\frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{x}^\mathsf{T} - \boldsymbol{\lambda}^\mathsf{T} A = \mathbf{0}^\mathsf{T},$$
$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = -(A\mathbf{x} - \mathbf{b}) = \mathbf{0}.$$

From the first equation we have $\mathbf{x} = \frac{1}{2}A^\mathsf{T}\boldsymbol{\lambda}$. We plug this into the second equation to get

$$\frac{1}{2}AA^\mathsf{T}\boldsymbol{\lambda} = \mathbf{b}.$$

If we assume that A is of full rank (so that AA^T is invertible), we can express $\boldsymbol{\lambda} = 2\left(AA^\mathsf{T}\right)^{-1}\mathbf{b}$, and therefore

$$\mathbf{x} = \frac{1}{2}A^\mathsf{T}\boldsymbol{\lambda} = A^\mathsf{T}\left(AA^\mathsf{T}\right)^{-1}\mathbf{b}.$$

In general, ie. for non-full rank matrices A, the solution is $\mathbf{x} = A^+\mathbf{b}$, where A^+ is the Moore–Penrose inverse of A.
Finally, the minimal value of f on that affine subspace is

$$f\left(A^\mathsf{T}\left(AA^\mathsf{T}\right)^{-1}\mathbf{b}\right) = \mathbf{b}^\mathsf{T}\left(AA^\mathsf{T}\right)^{-1}AA^\mathsf{T}\left(AA^\mathsf{T}\right)^{-1}\mathbf{b} = \mathbf{b}^\mathsf{T}\left(AA^\mathsf{T}\right)^{-1}\mathbf{b}$$

for the case when A is of full rank. (That -1 superscript needs to be replaced by a $+$ superscript for A which is not of full rank.)

(c) Let's consider the boundary first, ie. we minimize $f(\mathbf{x}) = \|\mathbf{x}\|^2$ with respect to $\|\mathbf{x} - \mathbf{p}\| = d$ and $A\mathbf{x} = \mathbf{b}$. The Lagrange function is

$$L(\mathbf{x}, \mu, \boldsymbol{\lambda}) = \mathbf{x}^\mathsf{T}\mathbf{x} - \mu(\|\mathbf{x} - \mathbf{p}\|^2 - d^2) - \boldsymbol{\lambda}^\mathsf{T}(A\mathbf{x} - \mathbf{b}).$$

Its stationary points are solutions of the system

$$\frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{x}^\mathsf{T} - 2\mu(\mathbf{x} - \mathbf{p})^\mathsf{T} - \boldsymbol{\lambda}^\mathsf{T} A = \mathbf{0}^\mathsf{T},$$
$$\frac{\partial L}{\partial \mu} = -(\|\mathbf{x} - \mathbf{p}\|^2 - d^2) = 0,$$
$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = -(A\mathbf{x} - \mathbf{b}) = \mathbf{0}.$$

Let's start: From the first equation we get

$$(2 - 2\mu)\mathbf{x} = A^\mathsf{T}\boldsymbol{\lambda} - 2\mu\mathbf{p} \quad \therefore \quad \mathbf{x} = \frac{1}{2 - 2\mu}\left(A^\mathsf{T}\boldsymbol{\lambda} - 2\mu\mathbf{p}\right).$$

Plugging this into the third equation we have

$$A\left(A^\mathsf{T}\boldsymbol{\lambda} - 2\mu\mathbf{p}\right) = (2 - 2\mu)\mathbf{x} \quad \therefore \quad AA^\mathsf{T}\boldsymbol{\lambda} = (2 - 2\mu)\mathbf{b} + 2\mu A\mathbf{p},$$

hence

$$\boldsymbol{\lambda} = \left(AA^\top\right)^{-1}((2-2\mu)\mathbf{b} + 2\mu A\mathbf{p}).$$

Now we plug this into the expression for $\mathbf{x}$ and obtain

$$\begin{aligned}
\mathbf{x} &= \frac{1}{2-2\mu}\left(A^\top\left(\left(AA^\top\right)^{-1}((2-2\mu)\mathbf{b} + 2\mu A\mathbf{p})\right) - 2\mu\mathbf{p}\right) \\
&= \frac{1}{2-2\mu}\left((2-2\mu)A^\top\left(AA^\top\right)^{-1}\mathbf{b} + 2\mu(A^\top A\mathbf{p} - \mathbf{p})\right) \\
&= A^\top\left(AA^\top\right)^{-1}\mathbf{b} + \frac{\mu}{1-\mu}(A^\top A - I)\mathbf{p} \\
&= A^\top\left(AA^\top\right)^{-1}\mathbf{b} + \alpha(A^\top A - I)\mathbf{p},
\end{aligned}$$

where we introduced $\alpha = \frac{\mu}{1-\mu}$ in the last line. Finally, we use the constraint $\|\mathbf{x}-\mathbf{p}\|^2 = d^2$. Plugging the above expression for $\mathbf{x}$ into it and rearranging we obtain a quadratic equation for α, namely

$$\begin{aligned}
\alpha^2\,\mathbf{p}^\top(A^\top A - I)\mathbf{p} + 2\alpha\left(\mathbf{p}^\top A^\top\mathbf{b} - \mathbf{p}^\top A^\top(AA^\top)^{-1}\mathbf{b} - \mathbf{p}^\top(A^\top A - I)\mathbf{p}\right) + \\
+ \mathbf{b}^\top(AA^\top)^{-1}\mathbf{b} - 2\mathbf{p}^\top A^\top(AA^\top)^{-1}\mathbf{b} + \mathbf{p}^\top\mathbf{p} - d^2 = 0
\end{aligned}$$

Solving this equation for α and plugging the solution in to our expression for $\mathbf{x}$, we finally obtain the points, at which extreme values of $f(\mathbf{x}) = \mathbf{x}^\top\mathbf{x}$ are attained. Conveniently, we leave that final task to the reader.