2

Linear Algebra

Exercises

2.1 We consider $(\mathbb{R}\backslash\{-1\}, \star)$, where

$$a \star b := ab + a + b, \qquad a, b \in \mathbb{R}\backslash\{-1\} \tag{2.1}$$

a. Show that $(\mathbb{R}\backslash\{-1\}, \star)$ is an Abelian group.
b. Solve

$$3 \star x \star x = 15$$

in the Abelian group $(\mathbb{R}\backslash\{-1\}, \star)$, where $\star$ is defined in (2.1).

a. First, we show that $\mathbb{R}\backslash\{-1\}$ is closed under $\star$: For all $a, b \in \mathbb{R}\backslash\{-1\}$:

$$a \star b = ab + a + b + 1 - 1 = \underbrace{(a+1)}_{\neq 0}\underbrace{(b+1)}_{\neq 0} - 1 \neq -1$$

$$\Rightarrow a \star b \in \mathbb{R}\backslash\{-1\}$$

Next, we show the group axioms

- **Associativity:** For all $a, b, c \in \mathbb{R}\backslash\{-1\}$:

$$\begin{aligned}(a \star b) \star c &= (ab + a + b) \star c \\ &= (ab + a + b)c + (ab + a + b) + c \\ &= abc + ac + bc + ab + a + b + c \\ &= a(bc + b + c) + a + (bc + b + c) \\ &= a \star (bc + b + c) \\ &= a \star (b \star c)\end{aligned}$$

- **Commutativity:**

$$\forall a, b \in \mathbb{R}\backslash\{-1\} : a \star b = ab + a + b = ba + b + a = b \star a$$

- **Neutral Element:** $n = 0$ is the neutral element since

$$\forall a \in \mathbb{R}\backslash\{-1\} : a \star 0 = a = 0 \star a$$

- **Inverse Element:** We need to find $\bar{a}$, such that $a \star \bar{a} = 0 = \bar{a} \star a$.

$$\begin{aligned}\bar{a} \star a = 0 &\iff \bar{a}a + a + \bar{a} = 0 \\ &\iff \bar{a}(a+1) = -a \\ &\overset{a \neq -1}{\iff} \bar{a} = -\frac{a}{a+1} = -1 + \frac{1}{a+1} \neq -1 \in \mathbb{R}\backslash\{-1\}\end{aligned}$$

b.

$$
\begin{aligned}
3 \star x \star x = 15 &\iff 3 \star (x^2 + x + x) = 15 \\
&\iff 3x^2 + 6x + 3 + x^2 + 2x = 15 \\
&\iff 4x^2 + 8x - 12 = 0 \\
&\iff (x-1)(x+3) = 0 \\
&\iff x \in \{-3, 1\}
\end{aligned}
$$

2.2 Let n be in $\mathbb{N}\backslash\{0\}$. Let k, x be in $\mathbb{Z}$. We define the congruence class $\bar{k}$ of the integer k as the set

$$
\begin{aligned}
\overline{k} &= \{x \in \mathbb{Z} \mid x - k = 0 \pmod{n}\} \\
&= \{x \in \mathbb{Z} \mid \exists a \in \mathbb{Z} \colon (x - k = n \cdot a)\} \,.
\end{aligned}
$$

We now define $\mathbb{Z}/n\mathbb{Z}$ (sometimes written $\mathbb{Z}_n$) as the set of all congruence classes modulo n. Euclidean division implies that this set is a finite set containing n elements:

$$\mathbb{Z}_n = \{\overline{0}, \overline{1}, \ldots, \overline{n-1}\}$$

For all $\overline{a}, \overline{b} \in \mathbb{Z}_n$, we define

$$\overline{a} \oplus \overline{b} := \overline{a+b}$$

a. Show that $(\mathbb{Z}_n, \oplus)$ is a group. Is it Abelian?

b. We now define another operation $\otimes$ for all $\overline{a}$ and $\overline{b}$ in $\mathbb{Z}_n$ as

$$\overline{a} \otimes \overline{b} = \overline{a \times b}\,, \tag{2.2}$$

where $a \times b$ represents the usual multiplication in $\mathbb{Z}$.
Let $n = 5$. Draw the times table of the elements of $\mathbb{Z}_5\backslash\{\overline{0}\}$ under $\otimes$, i.e., calculate the products $\overline{a} \otimes \overline{b}$ for all $\overline{a}$ and $\overline{b}$ in $\mathbb{Z}_5\backslash\{\overline{0}\}$.
Hence, show that $\mathbb{Z}_5\backslash\{\overline{0}\}$ is closed under $\otimes$ and possesses a neutral element for $\otimes$. Display the inverse of all elements in $\mathbb{Z}_5\backslash\{\overline{0}\}$ under $\otimes$. Conclude that $(\mathbb{Z}_5\backslash\{\overline{0}\}, \otimes)$ is an Abelian group.

c. Show that $(\mathbb{Z}_8\backslash\{\overline{0}\}, \otimes)$ is not a group.

d. We recall that the Bézout theorem states that two integers a and b are relatively prime (i.e., $gcd(a, b) = 1$) if and only if there exist two integers u and v such that $au + bv = 1$. Show that $(\mathbb{Z}_n\backslash\{\overline{0}\}, \otimes)$ is a group if and only if $n \in \mathbb{N}\backslash\{0\}$ is prime.

a. We show that the group axioms are satisfied:

- Closure: Let $\overline{a}, \overline{b}$ be in $\mathbb{Z}_n$. We have:

$$
\begin{aligned}
\overline{a} \oplus \overline{b} &= \overline{a+b} \\
&= \overline{(a+b) \mod n}
\end{aligned}
$$

by definition of the congruence class, and since $[(a+b) \bmod n] \in \{0, \ldots, n-1\}$, it follows that $\overline{a} \oplus \overline{b} \in \mathbb{Z}_n$. Thus, $\mathbb{Z}_n$ is closed under $\oplus$.

- Associativity: Let $\bar{c}$ be in $\mathbb{Z}_n$. We have:

$$(\bar{a} \oplus \bar{b}) \oplus \bar{c} = \overline{(a+b)} \oplus \bar{c} = \overline{(a+b)+c} = \overline{a+(b+c)}$$
$$= \bar{a} \oplus \overline{(b+c)} = \bar{a} \oplus (\bar{b} \oplus \bar{c})$$

  so that $\oplus$ is associative.
- Neutral element: We have

$$\bar{a} + \bar{0} = \overline{a+0} = \bar{a} = \bar{0} + \bar{a}$$

  so $\bar{0}$ is the neutral element for $\oplus$.
- Inverse element: We have

$$\bar{a} + \overline{(-a)} = \overline{a-a} = \bar{0} = \overline{(-a)} + \bar{a}$$

  and we know that $\overline{(-a)}$ is equal to $\overline{(-a) \bmod n}$ which belongs to $\mathbb{Z}_n$ and is thus the inverse of $\bar{a}$.
- Commutativity: Finally, the commutativity of $(\mathbb{Z}_n, \oplus)$ follows from that of $(\mathbb{Z}, +)$ since we have

$$\bar{a} \oplus \bar{b} = \overline{a+b} = \overline{b+a} = \bar{b} \oplus \bar{a}\,,$$

  which shows that $(\mathbb{Z}_n, \oplus)$ is an Abelian group.

b. Let us calculate the times table of $\mathbb{Z}_5 \backslash \{\bar{0}\}$ under $\otimes$:

$\otimes$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$
$\bar{1}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$
$\bar{2}$	$\bar{2}$	$\bar{4}$	$\bar{1}$	$\bar{3}$
$\bar{3}$	$\bar{3}$	$\bar{1}$	$\bar{4}$	$\bar{2}$
$\bar{4}$	$\bar{4}$	$\bar{3}$	$\bar{2}$	$\bar{1}$

We can notice that all the products are in $\mathbb{Z}_5 \backslash \{\bar{0}\}$, and that in particular, none of them is equal to $\bar{0}$. Thus, $\mathbb{Z}_5 \backslash \{\bar{0}\}$ is closed under $\otimes$. The neutral element is $\bar{1}$ and we have $(\bar{1})^{-1} = \bar{1}$, $(\bar{2})^{-1} = \bar{3}$, $(\bar{3})^{-1} = \bar{2}$, and $(\bar{4})^{-1} = \bar{4}$. Associativity and commutativity are straightforward and $(\mathbb{Z}_5 \backslash \{\bar{0}\}, \otimes)$ is an Abelian group.

c. The elements $\bar{2}$ and $\bar{4}$ belong to $\mathbb{Z}_8 \backslash \{\bar{0}\}$, but their product $\bar{2} \otimes \bar{4} = \bar{8} = \bar{0}$ does not. Thus, this set is not closed under $\otimes$ and is not a group.

d.
- Let us assume that n is not prime and can thus be written as a product $n = a \times b$ of two integers a and b in $\{2, \ldots, n-1\}$. Both elements $\bar{a}$ and $\bar{b}$ belong to $\mathbb{Z}_n \backslash \{\bar{0}\}$ but their product $\bar{a} \otimes \bar{b} = \bar{n} = \bar{0}$ does not. Thus, this set is not closed under $\otimes$ and $(\mathbb{Z}_n \backslash \{\bar{0}\}, \otimes)$ is not a group.
- Let n be a prime number. Let $\bar{a}$ and $\bar{b}$ be in $\mathbb{Z}_n \backslash \{\bar{0}\}$ with a and b in $\{1, \ldots, n-1\}$. As n is prime, we know that a is relatively prime to n, and so is b. Let us then take four integers u, v, u' and v', such that

$$au + nv = 1$$
$$bu' + nv' = 1\,.$$

  We thus have that $(au + nv)(bu' + nv') = 1$, which we can rewrite as

$$ab(uu') + n(auv' + vbu' + nvv') = 1$$

By virtue of the Bézout theorem, this implies that ab and n are relatively prime, which ensures that the product $\overline{a} \otimes \overline{b}$ is not equal to $\overline{0}$ and belongs to $\mathbb{Z}_n \backslash \{\overline{0}\}$, which is thus closed under $\otimes$.
The associativity and commutativity of $\otimes$ are straightforward, but we need to show that every element has an inverse. First, the neutral element is $\overline{1}$. Let us again consider an element $\overline{a}$ in $\mathbb{Z}_n \backslash \{\overline{0}\}$ with a in $\{1, \ldots, n-1\}$. As a and n are coprime, the Bézout theorem enables us to define two integers u and v such that

$$au + nv = 1\,, \tag{2.3}$$

which implies that $au = 1 - nv$ and thus

$$au = 1 \mod n\,, \tag{2.4}$$

which means that $\overline{a} \otimes \overline{u} = \overline{au} = \overline{1}$, or that $\overline{u}$ is the inverse of $\overline{a}$. Overall, $(\mathbb{Z}_n \backslash \{\overline{0}\}, \otimes)$ is an Abelian group. Note that the Bézout theorem ensures the existence of an inverse without yielding its explicit value, which is the purpose of the extended Euclidean algorithm.

2.3 Consider the set $\mathcal{G}$ of 3×3 matrices defined as follows:

$$\mathcal{G} = \left\{ \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3} \,\middle|\, x, y, z \in \mathbb{R} \right\}$$

We define $\cdot$ as the standard matrix multiplication.
Is $(\mathcal{G}, \cdot)$ a group? If yes, is it Abelian? Justify your answer.

- Closure: Let a, b, c, x, y and z be in $\mathbb{R}$ and let us define $\boldsymbol{A}$ and $\boldsymbol{B}$ in $\mathcal{G}$ as

$$\boldsymbol{A} = \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix}.$$

Then,

$$\boldsymbol{A} \cdot \boldsymbol{B} = \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & a + x & c + xb + z \\ 0 & 1 & b + y \\ 0 & 0 & 1 \end{bmatrix}.$$

Since $a + x$, $b + y$ and $c + xb + z$ are in $\mathbb{R}$ we have $\boldsymbol{A} \cdot \boldsymbol{B} \in \mathcal{G}$. Thus, $\mathcal{G}$ is closed under matrix multiplication.

- Associativity: Let α, β and γ be in $\mathbb{R}$ and let $\boldsymbol{C}$ in $\mathcal{G}$ be defined as

$$\boldsymbol{C} = \begin{bmatrix} 1 & \alpha & \gamma \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{bmatrix}.$$

It holds that

$$\begin{aligned}(\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{C} &= \begin{bmatrix} 1 & a + x & c + xb + z \\ 0 & 1 & b + y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \alpha & \gamma \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & \alpha + a + x & \gamma + \alpha\beta + x\beta + c + xb + z \\ 0 & 1 & \beta + b + y \\ 0 & 0 & 1 \end{bmatrix}.\end{aligned}$$

Similarly,

$$\boldsymbol{A} \cdot (\boldsymbol{B} \cdot \boldsymbol{C}) = \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & \alpha + a & \gamma + \alpha\beta + c \\ 0 & 1 & \beta + b \\ 0 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & \alpha + a + x & \gamma + \alpha\beta + x\beta + c + xb + z \\ 0 & 1 & \beta + b + y \\ 0 & 0 & 1 \end{bmatrix} = (\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{C} .$$

Therefore, $\cdot$ is associative.

- Neutral element: For all $\boldsymbol{A}$ in $\mathcal{G}$, we have: $\boldsymbol{I}_3 \cdot \boldsymbol{A} = \boldsymbol{A} = \boldsymbol{A} \cdot \boldsymbol{I}_3$ and thus $\boldsymbol{I}_3$ is the neutral element.
- Non-commutativity: We show that $\cdot$ is not commutative. Consider the matrices $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{G}$, where

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} , \qquad \boldsymbol{Y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} . \tag{2.5}$$

Then,

$$\boldsymbol{X} \cdot \boldsymbol{Y} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} ,$$
$$\boldsymbol{Y} \cdot \boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \neq \boldsymbol{X} \cdot \boldsymbol{Y} .$$

Therefore, $\cdot$ is not commutative.

- Inverse element: Let us look for a right inverse $\boldsymbol{A}_r^{-1}$ of $\boldsymbol{A}$. Such a matrix should satisfy $\boldsymbol{A}\boldsymbol{A}_r^{-1} = \boldsymbol{I}_3$. We thus solve the linear system $[\boldsymbol{A}|\boldsymbol{I}_3]$ that we transform into $[\boldsymbol{I}_3|\boldsymbol{A}_r^{-1}]$:

$$\left[\begin{array}{ccc|ccc} 1 & x & z & 1 & 0 & 0 \\ 0 & 1 & y & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} -zR_3 \\ -yR_3 \\ \\ \end{array}$$
$$\rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & x & 0 & 1 & 0 & -z \\ 0 & 1 & 0 & 0 & 1 & -y \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} -xR_2 \\ \\ \\ \end{array}$$
$$\rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -x & xy - z \\ 0 & 1 & 0 & 0 & 1 & -y \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] .$$

Therefore, we obtain the right inverse

$$\boldsymbol{A}_r^{-1} = \begin{bmatrix} 1 & -x & xy - z \\ 0 & 1 & -y \\ 0 & 0 & 1 \end{bmatrix} \in \mathcal{G}$$

Because of the uniqueness of the inverse element, if a left inverse $\boldsymbol{A}_l^{-1}$ exists, then it is equal to the right inverse. But as $\cdot$ is not commutative, we need to check manually that we also have $\boldsymbol{A}\boldsymbol{A}_r^{-1} = \boldsymbol{I}_3$, which we do next:

$$\boldsymbol{A}_r^{-1}\boldsymbol{A} = \begin{bmatrix} 1 & -x & -z+xy \\ 0 & 1 & -y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & x-x & z-xy-z+xy \\ 0 & 1 & y+y \\ 0 & 0 & 1 \end{bmatrix} = \boldsymbol{I}_3 \,.$$

Thus, every element of $\mathcal{G}$ has an inverse. Overall, $(\mathcal{G}, \cdot)$ is a non-Abelian group.

2.4 Compute the following matrix products, if possible:

a.

$$\begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

This matrix product is not defined. Highlight that the neighboring dimensions have to fit (i.e., $m \times n$ matrices need to be multiplied by $n \times p$ (from the right) or $k \times m$ matrices (from the left).)

b.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 5 \\ 10 & 9 & 11 \\ 16 & 15 & 17 \end{bmatrix}$$

c.

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 5 & 7 & 9 \\ 11 & 13 & 15 \\ 8 & 10 & 12 \end{bmatrix}$$

d.

$$\begin{bmatrix} 1 & 2 & 1 & 2 \\ 4 & 1 & -1 & -4 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 1 & -1 \\ 2 & 1 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ -21 & 2 \end{bmatrix}$$

e.

$$\begin{bmatrix} 0 & 3 \\ 1 & -1 \\ 2 & 1 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 2 \\ 4 & 1 & -1 & -4 \end{bmatrix} = \begin{bmatrix} 12 & 3 & -3 & -12 \\ -3 & 1 & 2 & 6 \\ 6 & 5 & 1 & 0 \\ 13 & 12 & 3 & 2 \end{bmatrix}$$

2.5 Find the set $\mathcal{S}$ of all solutions in $\boldsymbol{x}$ of the following inhomogeneous linear systems $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A}$ and $\boldsymbol{b}$ are defined as follows:

a.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 2 & 5 & -7 & -5 \\ 2 & -1 & 1 & 3 \\ 5 & 2 & -4 & 2 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ -2 \\ 4 \\ 6 \end{bmatrix}$$

We apply Gaussian elimination to the augmented matrix

$$\left[\begin{array}{cccc|c} 1 & 1 & -1 & -1 & 1 \\ 2 & 5 & -7 & -5 & -2 \\ 2 & -1 & 1 & 3 & 4 \\ 5 & 2 & -4 & 2 & 6 \end{array}\right] \begin{array}{l} \\ -2R_1 \\ -2R_1 \\ -5R_1 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{cccc|c} 1 & 1 & -1 & -1 & 1 \\ 0 & 3 & -5 & -3 & -4 \\ 0 & -3 & 3 & 5 & 2 \\ 0 & -3 & 1 & 7 & 1 \end{array}\right] \begin{array}{l} -\frac{1}{3}R_2 \\ \mid \cdot \frac{1}{3} \\ +R_2 \\ +R_2 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{cccc|c} 1 & 0 & \frac{2}{3} & 0 & \frac{7}{3} \\ 0 & 1 & -\frac{5}{3} & -1 & -\frac{4}{3} \\ 0 & 0 & -2 & 2 & -2 \\ 0 & 0 & -4 & 4 & -3 \end{array}\right] \begin{array}{l} \\ \\ \\ -2R_3 \end{array} \rightsquigarrow \left[\begin{array}{cccc|c} 1 & 0 & \frac{2}{3} & 0 & \frac{7}{3} \\ 0 & 1 & -\frac{5}{3} & -1 & -\frac{4}{3} \\ 0 & 0 & -2 & 2 & -2 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

The last row of the final linear system shows that the equation system has no solution and thus $S = \emptyset$.

b.

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 \\ 1 & 1 & 0 & -3 & 0 \\ 2 & -1 & 0 & 1 & -1 \\ -1 & 2 & 0 & -2 & -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 6 \\ 5 \\ -1 \end{bmatrix}$$

We start again by writing down the augmented matrix and apply Gaussian elimination to obtain the reduced row echelon form:

$$\left[\begin{array}{ccccc|c} 1 & -1 & 0 & 0 & 1 & 3 \\ 1 & 1 & 0 & -3 & 0 & 6 \\ 2 & -1 & 0 & 1 & -1 & 5 \\ -1 & 2 & 0 & -2 & -1 & -1 \end{array}\right] \begin{array}{l} \\ -R_1 \\ -2R_1 \\ +R_1 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccccc|c} 1 & -1 & 0 & 0 & 1 & 3 \\ 0 & 2 & 0 & -3 & -1 & 3 \\ 0 & 1 & 0 & 1 & -3 & -1 \\ 0 & 1 & 0 & -2 & 0 & 2 \end{array}\right] \begin{array}{l} +R_3 \\ -2R_3 \\ \text{swap with } R_2 \\ -R_3 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 1 & -2 & 2 \\ 0 & 1 & 0 & 1 & -3 & -1 \\ 0 & 0 & 0 & -5 & 5 & 5 \\ 0 & 0 & 0 & -3 & 3 & 3 \end{array}\right] \begin{array}{l} \\ \\ \cdot(-\frac{1}{5}) \\ +\frac{5}{3}R_3 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 1 & -2 & 2 \\ 0 & 1 & 0 & 1 & -3 & -1 \\ 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \begin{array}{l} -R_3 \\ -R_3 \\ \\ \\ \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & -1 & 3 \\ 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{array}\right]$$

From the reduced row echelon form we apply the “Minus-1 Trick” in order to get the following system:

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & -1 & 3 \\ 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{array}\right]$$

The right-hand side of the system yields us a particular solution while the columns corresponding to the -1 pivots are the directions of the solution space. We then obtain the set of all possible solutions as

$$\mathcal{S} := \left\{ \boldsymbol{x} \in \mathbb{R}^6 : \boldsymbol{x} = \begin{bmatrix} 3 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} -1 \\ -2 \\ 0 \\ -1 \\ -1 \end{bmatrix} \middle| \lambda_1, \lambda_2 \in \mathbb{R} \right\} .$$

2.6 Using Gaussian elimination, find all solutions of the inhomogeneous equation system $\boldsymbol{Ax} = \boldsymbol{b}$ with

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} .$$

We start by determining the reduced row echelon form of the augmented matrix $[\boldsymbol{A}|\boldsymbol{b}]$.

$$\begin{array}{l} \left[\begin{array}{cccccc|c} 0 & 1 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \end{array}\right] \begin{array}{l} \\ \\ -R_1 \end{array} \\ \rightsquigarrow \left[\begin{array}{cccccc|c} 0 & 1 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 & -1 \end{array}\right] \begin{array}{l} +R_3 \\ +R_3 \\ \cdot(-1) \end{array} \\ \rightsquigarrow \left[\begin{array}{cccccc|c} 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{array}\right] \end{array}$$

This augmented matrix is now in reduced row echelon form. Applying the "Minus-1 trick" gives us the following augmented matrix:

$$\left[\begin{array}{cccccc|c} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{array}\right]$$

The right-hand side of this augmented matrix gives us a particular solution while the columns corresponding to the -1 pivots are the span the solution space S. Therefore, we obtain the general solution

$$S = \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + \underbrace{\lambda_1 \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ -1 \\ -1 \end{bmatrix}}_{\text{solves } \boldsymbol{Ax}=\boldsymbol{0}} \;\middle|\; \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R} \right\}$$

Now, we find a particular solution that solves the inhomogeneous equation system. From the RREF, we see that $x_2 = 1, x_4 = -2, x_5 = 1 + x_6$ and $x_1, x_3, x_6 \in \mathbb{R}$ are free variables (they correspond to variables that belong to the non-pivot columns of the augmented matrix). Therefore, a particular solution is

$$\boldsymbol{x}_p = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} \in \mathbb{R}^6 .$$

The general solution adds solutions from the homogeneous equation system $\boldsymbol{Ax} = \boldsymbol{0}$. We can use the RREF of the augmented system to read out these solutions by using the Minus-1 Trick. Padding the RREF with rows containing -1 on the diagonal gives us

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

This yields the general solution

$$\boldsymbol{x}_p + \underbrace{\lambda_1 \begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ -1 \\ -1 \end{bmatrix}}_{\text{solves } \boldsymbol{Ax}=\boldsymbol{0}}, \quad \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R} .$$

2.7 Find all solutions in $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3$ of the equation system $\boldsymbol{Ax} = 12\boldsymbol{x}$, where

$$\boldsymbol{A} = \begin{bmatrix} 6 & 4 & 3 \\ 6 & 0 & 9 \\ 0 & 8 & 0 \end{bmatrix}$$

and $\sum_{i=1}^{3} x_i = 1$.

We start by rephrasing the problem into solving a homogeneous system of linear equations. Let $\boldsymbol{x}$ be in $\mathbb{R}^3$. We notice that $\boldsymbol{A}\boldsymbol{x} = 12\boldsymbol{x}$ is equivalent to $(\boldsymbol{A} - 12\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$, which can be rewritten as the homogeneous system $\tilde{\boldsymbol{A}}\boldsymbol{x} = \boldsymbol{0}$, where we define

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} -6 & 4 & 3 \\ 6 & -12 & 9 \\ 0 & 8 & -12 \end{bmatrix}.$$

The constraint $\sum_{i=1}^{3} x_i = 1$ can be transcribed as a fourth equation, which leads us to consider the following linear system, which we bring to reduced row echelon form:

$$\left[\begin{array}{ccc|c} -6 & 4 & 3 & 0 \\ 6 & -12 & 9 & 0 \\ 0 & 8 & -12 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right] \begin{array}{l} +R_2 \\ \cdot\frac{1}{3} \\ \cdot\frac{1}{4} \\ \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 0 & -8 & 12 & 0 \\ 2 & -4 & 3 & 0 \\ 0 & 2 & -3 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right] \begin{array}{l} +4R_3 \\ +2R_3 \\ \\ \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 2 & 0 & -3 & 0 \\ 0 & 2 & -3 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right] \begin{array}{l} \\ \cdot\frac{1}{2} \\ \cdot\frac{1}{2} \\ \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -\frac{3}{2} & 0 \\ 0 & 1 & -\frac{3}{2} & 0 \\ 1 & 1 & 1 & 1 \end{array}\right] \begin{array}{l} \\ \\ -R_1 - R_2 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -\frac{3}{2} & 0 \\ 0 & 1 & -\frac{3}{2} & 0 \\ 0 & 0 & 4 & 1 \end{array}\right] \begin{array}{l} +(\frac{3}{8})R_3 \\ +(\frac{3}{8})R_3 \\ \cdot\frac{1}{4} \end{array} \rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{3}{8} \\ 0 & 1 & 0 & \frac{3}{8} \\ 0 & 0 & 1 & \frac{1}{4} \end{array}\right]$$

Therefore, we obtain the unique solution

$$\boldsymbol{x} = \begin{bmatrix} \frac{3}{8} \\ \frac{3}{8} \\ \frac{1}{4} \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 3 \\ 3 \\ 2 \end{bmatrix}.$$

2.8 Determine the inverses of the following matrices if possible:

a.

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}$$

To determine the inverse of a matrix, we start with the augmented matrix $[\boldsymbol{A} \,|\, \boldsymbol{I}]$ and transform it into $[\boldsymbol{I} \,|\, \boldsymbol{B}]$, where $\boldsymbol{B}$ turns out to be $\boldsymbol{A}^{-1}$:

$$\left[\begin{array}{ccc|ccc} 2 & 3 & 4 & 1 & 0 & 0 \\ 3 & 4 & 5 & 0 & 1 & 0 \\ 4 & 5 & 6 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} \\ -\frac{3}{2}R_1 \\ -2R_1 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|ccc} 2 & 3 & 4 & 1 & 0 & 0 \\ 0 & -\frac{1}{2} & -1 & -\frac{3}{2} & 1 & 0 \\ 0 & -1 & -2 & -2 & 0 & 1 \end{array}\right] \begin{array}{l} \\ -\frac{1}{2}R_3 \\ \cdot(-1) \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|ccc} 2 & 3 & 4 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 1 & 2 & 2 & 0 & -1 \end{array}\right].$$

Here, we see that this system of linear equations is not solvable. Therefore, the inverse does not exist.

b.

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

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} \\ \\ -R_1 \\ -R_1 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{cccc|cccc} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} \\ -R_4 \\ -R_4 \\ \text{swap with } R_2 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{cccc|cccc} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & -1 \end{array}\right] \begin{array}{l} -R_4 \\ \\ +R_4 \\ \text{swap with } R_3 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & -2 \end{array}\right]$$

Therefore,

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

2.9 Which of the following sets are subspaces of $\mathbb{R}^3$?

a. $A = \{(\lambda, \lambda + \mu^3, \lambda - \mu^3) \mid \lambda, \mu \in \mathbb{R}\}$
b. $B = \{(\lambda^2, -\lambda^2, 0) \mid \lambda \in \mathbb{R}\}$
c. Let γ be in $\mathbb{R}$.
$C = \{(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 \mid \xi_1 - 2\xi_2 + 3\xi_3 = \gamma\}$
d. $D = \{(\xi_1, \xi_2, \xi_3) \in \mathbb{R}^3 \mid \xi_2 \in \mathbb{Z}\}$

As a reminder: Let V be a vector space. $U \subseteq V$ is a subspace if

1. $U \neq \emptyset$. In particular, $\mathbf{0} \in U$.
2. $\forall \boldsymbol{a}, \boldsymbol{b} \in U : \boldsymbol{a} + \boldsymbol{b} \in U$ Closure with respect to the inner operation
3. $\forall \boldsymbol{a} \in U, \lambda \in \mathbb{R} : \lambda \boldsymbol{a} \in U$ Closure with respect to the outer operation

The standard vector space properties (Abelian group, distributivity, associativity and neutral element) do not have to be shown because they are inherited from the vector space $(\mathbb{R}^3, +, \cdot)$.
Let us now have a look at the sets A, B, C, D.

a. 1. We have that $(0, 0, 0) \in A$ for $\lambda = 0 = \mu$.
2. Let $a = (\lambda_1, \lambda_1 + \mu_1^3, \lambda_1 - \mu_1^3)$ and $b = (\lambda_2, \lambda_2 + \mu_2^3, \lambda_2 - \mu_2^3)$ be two elements of A, where $\lambda_1, \mu_1, \lambda_2, \mu_2 \in \mathbb{R}$. Then,

$$\begin{aligned} a + b &= (\lambda_1, \lambda_1 + \mu_1^3, \lambda_1 - \mu_1^3) + (\lambda_2, \lambda_2 + \mu_2^3, \lambda_2 - \mu_2^3) \\ &= (\lambda_1 + \lambda_2, \lambda_1 + \mu_1^3 + \lambda_2 + \mu_2^3, \lambda_1 - \mu_1^3 + \lambda_2 - \mu_2^3) \\ &= (\lambda_1 + \lambda_2, (\lambda_1 + \lambda_2) + (\mu_1^3 + \mu_2^3), (\lambda_1 + \lambda_2) - (\mu_1^3 + \mu_2^3)), \end{aligned}$$

which belongs to A.
3. Let α be in $\mathbb{R}$. Then,

$$\alpha(\lambda, \lambda + \mu^3, \lambda - \mu^3) = (\alpha\lambda, \alpha\lambda + \alpha\mu^3, \alpha\lambda - \alpha\mu^3) \in A.$$

Therefore, A is a subspace of $\mathbb{R}^3$.

b. The vector $(1, -1, 0)$ belongs to B, but $(-1) \cdot (1, -1, 0) = (-1, 1, 0)$ does not. Thus, B is not closed under scalar multiplication and is not a subspace of $\mathbb{R}^3$.

c. Let $\boldsymbol{A} \in \mathbb{R}^{1\times 3}$ be defined as $\boldsymbol{A} = [1, -2, 3]$. The set C can be written as:

$$C = \{\boldsymbol{x} \in \mathbb{R}^3 \mid \boldsymbol{A}\boldsymbol{x} = \gamma\} .$$

We can first notice that $\mathbf{0}$ belongs to B only if $\gamma = 0$ since $\boldsymbol{A}\mathbf{=0}$. Let thus consider $\gamma = 0$ and ask whether C is a subspace of $\mathbb{R}^3$. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be in C. We know that $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$ and $\boldsymbol{A}\boldsymbol{y} = \mathbf{0}$, so that

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

Therefore, $\boldsymbol{x} + \boldsymbol{y}$ belongs to C. Let λ be in $\mathbb{R}$. Similarly,

$$\boldsymbol{A}(\lambda\boldsymbol{x}) = \lambda(\boldsymbol{A}\boldsymbol{x}) = \lambda\mathbf{0} = \mathbf{0}$$

Therefore, C is closed under scalar multiplication, and thus is a subspace of $\mathbb{R}^3$ if (and only if) $\gamma = 0$.

d. The vector $(0, 1, 0)$ belongs to D but $\pi(0, 1, 0)$ does not and thus D is not a subspace of $\mathbb{R}^3$.

2.10 Are the following sets of vectors linearly independent?

a.

$$\boldsymbol{x}_1 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} , \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} , \quad \boldsymbol{x}_3 = \begin{bmatrix} 3 \\ -3 \\ 8 \end{bmatrix}$$

To determine whether these vectors are linearly independent, we check if the $\mathbf{0}$-vector can be non-trivially represented as a linear combination of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_3$. Therefore, we try to solve the homogeneous linear equation system $\sum_{i=1}^{3} \lambda_i \boldsymbol{x}_i = \mathbf{0}$ for $\lambda_i \in \mathbb{R}$. We use Gaussian elimination to solve $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$ with

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 3 \\ -1 & 1 & -3 \\ 3 & -2 & 8 \end{bmatrix} ,$$

which leads to the reduced row echelon form

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

This means that $\boldsymbol{A}$ is rank deficient/singular and, therefore, the three vectors are linearly dependent. For example, with $\lambda_1 = 2, \lambda_2 = -1, \lambda_3 = -1$ we have a non-trivial linear combination $\sum_{i=1}^{3} \lambda_i \boldsymbol{x}_i = \mathbf{0}$.

b.

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} , \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} , \quad \boldsymbol{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Here, we are looking at the distribution of 0s in the vectors. $\boldsymbol{x}_1$ is the only vector whose third component is non-zero. Therefore, λ_1 must be 0. Similarly, λ_2 must be 0 because of the second component (already conditioning on $\lambda_1 = 0$). And finally, $\lambda_3 = 0$ as well. Therefore, the three vectors are linearly independent.
An alternative solution, using Gaussian elimination, is possible and would lead to the same conclusion.

2.11 Write

$$\boldsymbol{y} = \begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix}$$

as linear combination of

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \boldsymbol{x}_3 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

We are looking for $\lambda_1, \ldots, \lambda_3 \in \mathbb{R}$, such that $\sum_{i=1}^{3} \lambda_i \boldsymbol{x}_i = \boldsymbol{y}$. Therefore, we need to solve the inhomogeneous linear equation system

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 1 & 2 & -1 & -2 \\ 1 & 3 & 1 & 5 \end{array}\right]$$

Using Gaussian elimination, we obtain $\lambda_1 = -6, \lambda_2 = 3, \lambda_3 = 2$.

2.12 Consider two subspaces of $\mathbb{R}^4$:

$$U_1 = \operatorname{span}[\begin{bmatrix} 1 \\ 1 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}], \quad U_2 = \operatorname{span}[\begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 6 \\ -2 \\ -1 \end{bmatrix}].$$

Determine a basis of $U_1 \cap U_2$.
We start by checking whether there the vectors in the generating sets of U_1 (and U_2) are linearly dependent. Thereby, we can determine bases of U_1 and U_2, which will make the following computations simpler.
We start with U_1. To see whether the three vectors are linearly dependent, we need to find a linear combination of these vectors that allows a non-trivial representation of $\mathbf{0}$, i.e., $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$, such that

$$\lambda_1 \begin{bmatrix} 1 \\ 1 \\ -3 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

We see that necessarily: $\lambda_3 = -3\lambda_1$ (otherwise, the third component can never be 0). With this, we get

$$\lambda_1 \begin{bmatrix} 1+3 \\ 1-3 \\ -3+3 \\ 1-3 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\iff \lambda_1 \begin{bmatrix} 4 \\ -2 \\ 0 \\ -2 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and, therefore, $\lambda_2 = -2\lambda_1$. This means that there exists a non-trivial linear combination of $\mathbf{0}$ using spanning vectors of U_1, for example: $\lambda_1 = 1$, $\lambda_2 = -2$ and $\lambda_3 = -3$. Therefore, not all vectors in the generating set of U_1 are necessary, such that U_1 can be more compactly represented as

$$U_1 = \mathrm{span}[\begin{bmatrix} 1 \\ 1 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix}].$$

Now, we see whether the generating set of U_2 is also a basis. We try again whether we can find a non-trivial linear combination of $\mathbf{0}$ using the spanning vectors of U_2, i.e., a triple $(\alpha_1, \alpha_2, \alpha_3) \in \mathbb{R}^3$ such that

$$\alpha_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} -3 \\ 6 \\ -2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Here, we see that necessarily $\alpha_1 = \alpha_3$. Then, $\alpha_2 = 2\alpha_1$ gives a non-trivial representation of $\mathbf{0}$, and the three vectors are linearly dependent. However, any two of them are linearly independent, and we choose the first two vectors of the generating set as a basis of U_2, such that

$$U_2 = \mathrm{span}[\begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix}].$$

Now, we determine $U_1 \cap U_2$. Let $\boldsymbol{x}$ be in $\mathbb{R}^4$. Then,

$$\begin{aligned}
\boldsymbol{x} \in U_1 \cap U_2 &\iff \boldsymbol{x} \in U_1 \wedge \boldsymbol{x} \in U_2 \\
&\iff \exists \lambda_1, \lambda_2, \alpha_1, \alpha_2 \in \mathbb{R} : \left(\boldsymbol{x} = \alpha_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right) \\
&\qquad \wedge \left(\boldsymbol{x} = \lambda_1 \begin{bmatrix} 1 \\ 1 \\ -3 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix} \right) \\
&\iff \exists \lambda_1, \lambda_2, \alpha_1, \alpha_2 \in \mathbb{R} : \left(\boldsymbol{x} = \alpha_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right)
\end{aligned}$$

$$\wedge \left(\lambda_1 \begin{bmatrix} 1 \\ 1 \\ -3 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix} = \alpha_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right)$$

A general approach is to use Gaussian elimination to solve for either λ_1, λ_2 or α_1, α_2. In this particular case, we can find the solution by careful inspection: From the third component, we see that we need $-3\lambda_1 = 2\alpha_1$ and thus $\alpha_1 = -\frac{3}{2}\lambda_1$. Then:

$$\begin{aligned} \boldsymbol{x} \in U_1 \cap U_2 \iff \exists \lambda_1, \lambda_2, \alpha_2 \in \mathbb{R} \colon & \left(\boldsymbol{x} = -\tfrac{3}{2}\lambda_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right) \\ & \wedge \left(\lambda_1 \begin{bmatrix} 1 \\ 1 \\ -3 \\ 1 \end{bmatrix} + \tfrac{3}{2}\lambda_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix} = \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right) \\ \iff \exists \lambda_1, \lambda_2, \alpha_2 \in \mathbb{R} \colon & \left(\boldsymbol{x} = -\tfrac{3}{2}\lambda_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right) \\ & \wedge \left(\lambda_1 \begin{bmatrix} -\frac{1}{2} \\ -2 \\ 0 \\ \frac{5}{2} \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix} = \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right) \end{aligned}$$

The last component requires that $\lambda_2 = \frac{5}{2}\lambda_1$. Therefore,

$$\begin{aligned} \boldsymbol{x} \in U_1 \cap U_2 \iff \exists \lambda_1, \alpha_2 \in \mathbb{R} \colon & \left(\boldsymbol{x} = -\tfrac{3}{2}\lambda_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right) \\ & \wedge \left(\lambda_1 \begin{bmatrix} \frac{9}{2} \\ -\frac{9}{2} \\ 0 \\ 0 \end{bmatrix} = \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right) \\ \iff \exists \lambda_1, \alpha_2 \in \mathbb{R} \colon & \left(\boldsymbol{x} = -\tfrac{3}{2}\lambda_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right) \wedge (\alpha_2 = \tfrac{9}{4}\lambda_1) \\ \iff \exists \lambda_1 \in \mathbb{R} \colon & \left(\boldsymbol{x} = -\tfrac{3}{2}\lambda_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + \tfrac{9}{4}\lambda_1 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right) \end{aligned}$$

$$\iff \exists \lambda_1 \in \mathbb{R}\colon \left(\boldsymbol{x} = -6\lambda_1 \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix} + 9\lambda_1 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \right) \qquad \text{(mutiplied by 4)}$$

$$\iff \exists \lambda_1 \in \mathbb{R}\colon \left(\boldsymbol{x} = \lambda_1 \begin{bmatrix} 24 \\ -6 \\ -12 \\ -6 \end{bmatrix} \right)$$

$$\iff \exists \lambda_1 \in \mathbb{R}\colon \left(\boldsymbol{x} = \lambda_1 \begin{bmatrix} 4 \\ -1 \\ -2 \\ -1 \end{bmatrix} \right)$$

Thus, we have

$$U_1 \cap U_2 = \left\{ \lambda_1 \begin{bmatrix} 4 \\ -1 \\ -2 \\ -1 \end{bmatrix} \middle| \lambda_1 \in \mathbb{R} \right\} = \mathrm{span}[\begin{bmatrix} 4 \\ -1 \\ -2 \\ -1 \end{bmatrix},]$$

i.e., we obtain vector space spanned by $[4, -1, -2, -1]^\top$.

2.13 Consider two subspaces U_1 and U_2, where U_1 is the solution space of the homogeneous equation system $\boldsymbol{A}_1\boldsymbol{x} = \boldsymbol{0}$ and U_2 is the solution space of the homogeneous equation system $\boldsymbol{A}_2\boldsymbol{x} = \boldsymbol{0}$ with

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -2 & -1 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} 3 & -3 & 0 \\ 1 & 2 & 3 \\ 7 & -5 & 2 \\ 3 & -1 & 2 \end{bmatrix}.$$

a. Determine the dimension of U_1, U_2.

We determine U_1 by computing the reduced row echelon form of $\boldsymbol{A}_1$ as

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

which gives us

$$U_1 = \mathrm{span}[\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}].$$

Therefore, $\dim(U_1) = 1$. Similarly, we determine U_2 by computing the reduced row echelon form of $\boldsymbol{A}_2$ as

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

which gives us

$$U_2 = \mathrm{span}[\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}].$$

Therefore, $\dim(U_2) = 1$.

b. Determine bases of U_1 and U_2.
The basis vector that spans both U_1 and U_2 is

$$\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

c. Determine a basis of $U_1 \cap U_2$.
Since both U_1 and U_2 are spanned by the same basis vector, it must be that $U_1 = U_2$, and the desired basis is

$$U_1 \cap U_2 = U_1 = U_2 = \mathrm{span}[\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}].$$

2.14 Consider two subspaces U_1 and U_2, where U_1 is spanned by the columns of $\boldsymbol{A}_1$ and U_2 is spanned by the columns of $\boldsymbol{A}_2$ with

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -2 & -1 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} 3 & -3 & 0 \\ 1 & 2 & 3 \\ 7 & -5 & 2 \\ 3 & -1 & 2 \end{bmatrix}.$$

a. Determine the dimension of U_1, U_2
We start by noting that $U_1, U_2 \subseteq \mathbb{R}^4$ since we are interested in the space spanned by the columns of the corresponding matrices. Looking at $\boldsymbol{A}_1$, we see that $-\boldsymbol{d}_1 + \boldsymbol{d}_3 = \boldsymbol{d}_2$, where $\boldsymbol{d}_i$ are the columns of $\boldsymbol{A}_1$. This means that the second column can be expressed as a linear combination of $\boldsymbol{d}_1$ and $\boldsymbol{d}_3$. $\boldsymbol{d}_1$ and $\boldsymbol{d}_3$ are linearly independent, i.e., $\dim(U_1) = 2$.
Similarly, for $\boldsymbol{A}_2$, we see that the third column is the sum of the first two columns, and again we arrive at $\dim(U_2) = 2$.
Alternatively, we can use Gaussian elimination to determine a set of linearly independent columns in both matrices.

b. Determine bases of U_1 and U_2
A basis B of U_1 is given by the first two columns of $\boldsymbol{A}_1$ (any pair of columns would be fine), which are independent. A basis C of U_2 is given by the second and third columns of $\boldsymbol{A}_2$ (again, any pair of columns would be a basis), such that

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} \right\}, \quad C = \left\{ \begin{bmatrix} -3 \\ 2 \\ -5 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 2 \\ 2 \end{bmatrix} \right\}$$

c. Determine a basis of $U_1 \cap U_2$

Let us call $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{c}_1$ and $\boldsymbol{c}_2$ the vectors of the bases B and C such that $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2\}$ and $C = \{\boldsymbol{c}_1, \boldsymbol{c}_2\}$. Let $\boldsymbol{x}$ be in $\mathbb{R}^4$. Then,

$$
\begin{aligned}
\boldsymbol{x} \in U_1 \cap U_2 &\iff \exists \lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \mathbb{R} \colon (\boldsymbol{x} = \lambda_1 \boldsymbol{b}_1 + \lambda_2 \boldsymbol{b}_2) \wedge (\boldsymbol{x} = \lambda_3 \boldsymbol{c}_1 + \lambda_4 \boldsymbol{c}_2) \\
&\iff \exists \lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \mathbb{R} \colon (\boldsymbol{x} = \lambda_1 \boldsymbol{b}_1 + \lambda_2 \boldsymbol{b}_2) \\
&\qquad \wedge (\lambda_1 \boldsymbol{b}_1 + \lambda_2 \boldsymbol{b}_2 = \lambda_3 \boldsymbol{c}_1 + \lambda_4 \boldsymbol{c}_2) \\
&\iff \exists \lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \mathbb{R} \colon (\boldsymbol{x} = \lambda_1 \boldsymbol{b}_1 + \lambda_2 \boldsymbol{b}_2) \\
&\qquad \wedge (\lambda_1 \boldsymbol{b}_1 + \lambda_2 \boldsymbol{b}_2 - \lambda_3 \boldsymbol{c}_1 - \lambda_4 \boldsymbol{c}_2 = \boldsymbol{0})
\end{aligned}
$$

Let $\boldsymbol{\lambda} := [\lambda_1, \lambda_2, \lambda_3, \lambda_4]^\top$. The last equation of the system can be written as the linear system $\boldsymbol{A}\boldsymbol{\lambda} = \boldsymbol{0}$, where we define the matrix $\boldsymbol{A}$ as the concatenation of the column vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, -\boldsymbol{c}_1$ and $-\boldsymbol{c}_2$.

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

We solve this homogeneous linear system using Gaussian elimination.

$$
\begin{aligned}
&\begin{bmatrix} 1 & 0 & 3 & 0 \\ 1 & -2 & -2 & -3 \\ 2 & 1 & 5 & -2 \\ 1 & 0 & 1 & -2 \end{bmatrix} \begin{array}{l} \\ -R_1 \\ -2R_1 \\ -R_1 \end{array} \\
\rightsquigarrow &\begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & -2 & -5 & -3 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & -2 & -2 \end{bmatrix} \begin{array}{l} \\ +2R_3 \\ \text{swap with } R_2 \\ \cdot(-\frac{1}{2}) \end{array} \\
\rightsquigarrow &\begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & -7 & -7 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{array}{l} -3R_4 \\ +R_4 \\ +7R_4 \\ \text{swap with } R_3 \end{array} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}
\end{aligned}
$$

From the reduced row echelon form we find that the set

$$
S := \mathrm{span}[\begin{bmatrix} -3 \\ -1 \\ 1 \\ -1 \end{bmatrix}]
$$

describes the solution space of the system of equations in $\boldsymbol{\lambda}$.

We can now resume our equivalence derivation and replace the homogeneous system with its solution space. It holds

$$
\begin{aligned}
\boldsymbol{x} \in U_1 \cap U_2 &\iff \exists \lambda_1, \lambda_2, \lambda_3, \lambda_4, \alpha \in \mathbb{R} \colon (\boldsymbol{x} = \lambda_1 \boldsymbol{b}_1 + \lambda_2 \boldsymbol{b}_2) \\
&\qquad \wedge ([\lambda_1, \lambda_2, \lambda_3, \lambda_4]^\top = \alpha[-3, -1, 1, -1]^\top) \\
&\iff \exists \alpha \in \mathbb{R} \colon \boldsymbol{x} = -3\alpha \boldsymbol{b}_1 - \alpha \boldsymbol{b}_2 \\
&\iff \exists \alpha \in \mathbb{R} \colon \boldsymbol{x} = \alpha[-3, -1, -7, -3]^\top
\end{aligned}
$$

Finally,

$$U_1 \cap U_2 = \operatorname{span}[\begin{bmatrix} -3 \\ -1 \\ -7 \\ -3 \end{bmatrix}].$$

Alternatively, we could have expressed the solutions of $\boldsymbol{x}$ in terms of $\boldsymbol{b}_1$ and $\boldsymbol{c}_2$ with the condition on $\boldsymbol{\lambda}$ being $\exists \alpha \in \mathbb{R}\colon (\lambda_3 = \alpha) \wedge (\lambda_4 = -\alpha)$ to obtain $[3, 1, 7, 3]^\top$.

2.15 Let $F = \{(x, y, z) \in \mathbb{R}^3 \mid x+y-z = 0\}$ and $G = \{(a-b, a+b, a-3b) \mid a, b \in \mathbb{R}\}$.

a. Show that F and G are subspaces of $\mathbb{R}^3$.
We have $(0, 0, 0) \in F$ since $0 + 0 - 0 = 0$.
Let $\boldsymbol{a} = (x, y, z) \in \mathbb{R}^3$ and $\boldsymbol{b} = (x', y', z') \in \mathbb{R}^3$ be two elements of F. We have: $(x + y - z) + (x' + y' - z') = 0 + 0 = 0$ so $\boldsymbol{a} + \boldsymbol{b} \in F$.
Let $\lambda \in \mathbb{R}$. We also have: $\lambda x + \lambda y - \lambda z = \lambda 0 = 0$ so $\lambda \boldsymbol{a} \in F$ and thus F is a subspace of $\mathbb{R}^3$.
Similarly, we have $(0, 0, 0) \in G$ by setting a and b to 0.
Let a, b, a' and b' be in $\mathbb{R}$ and let $\boldsymbol{x} = (a - b, a + b, a - 3b)$ and $\boldsymbol{y} = (a' - b', a' + b', a' - 3b')$ be two elements of G. We have $\boldsymbol{x} + \boldsymbol{y} = ((a + a') - (b + b'), (a + a') + (b + b'), (a + a') - 3(b + b'))$ and $(a + a', b + b') \in \mathbb{R}^2$ so $\boldsymbol{x} + \boldsymbol{y} \in G$.
Let λ be in $\mathbb{R}$. We have $(\lambda a, \lambda b) \in \mathbb{R}^2$ so $\lambda \boldsymbol{x} \in G$ and thus G is a subspace of $\mathbb{R}^3$.

b. Calculate $F \cap G$ without resorting to any basis vector.
Combining both constraints, we have:

$$\begin{aligned}
F \cap G &= \{(a - b, a + b, a - 3b) \mid (a, b \in \mathbb{R}) \wedge [(a - b) + (a + b) - (a - 3b) = 0]\} \\
&= \{(a - b, a + b, a - 3b) \mid (a, b \in \mathbb{R}) \wedge (a = -3b)\} \\
&= \{(-4b, -2b, -6b) \mid b \in \mathbb{R}\} \\
&= \{(2b, b, 3b) \mid b \in \mathbb{R}\} \\
&= \operatorname{span}[\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}]
\end{aligned}$$

c. Find one basis for F and one for G, calculate $F \cap G$ using the basis vectors previously found and check your result with the previous question.
We can see that F is a subset of $\mathbb{R}^3$ with one linear constraint. It thus has dimension 2, and it suffices to find two independent vectors in F to construct a basis. By setting $(x, y) = (1, 0)$ and $(x, y) = (0, 1)$ successively, we obtain the following basis for F:

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Let us consider the set G. We introduce $u, v \in \mathbb{R}$ and perform the following variable substitutions: $u := a + b$ and $v := a - b$. Note that then

$a = (u+v)/2$ and $b = (u-v)/2$ and thus $a - 3b = 2v - u$, so that G can be written as

$$G = \{(v, u, 2v - u) \mid u, v \in \mathbb{R}\}.$$

The dimension of G is clearly 2 and a basis can be found by choosing two independent vectors of G, e.g.,

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\}$$

Let us now find $F \cap G$. Let $\boldsymbol{x} \in \mathbb{R}^3$. It holds that

$$\begin{aligned}
\boldsymbol{x} \in F \cap G \iff & \exists \lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{R} \colon \left(\boldsymbol{x} = \lambda_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right) \\
& \wedge \left(\boldsymbol{x} = \mu_1 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + \mu_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right) \\
\iff & \exists \lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{R} \colon \left(\boldsymbol{x} = \lambda_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right) \\
& \wedge \left(\lambda_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + \mu_1 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + \mu_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \boldsymbol{0} \right)
\end{aligned}$$

Note that for simplicity purposes, we have not reversed the sign of the coefficients for μ_1 and μ_2, which we can do since we could replace μ_1 by $-\mu_1$.

The latter equation is a linear system in $[\lambda_1, \lambda_2, \mu_1, \mu_2]^\top$ that we solve next.

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 2 & -1 \end{bmatrix} (\cdots) \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

The solution space for $(\lambda_1, \lambda_2, \mu_1, \mu_2)$ is therefore

$$\operatorname{span}[\begin{bmatrix} 2 \\ 1 \\ -2 \\ -1 \end{bmatrix}],$$

and we can resume our equivalence

$$\begin{aligned}
\boldsymbol{x} \in F \cap G \iff & \exists \alpha \in \mathbb{R} \colon \boldsymbol{x} = 2\alpha \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \\
\iff & \exists \alpha \in \mathbb{R} \colon \boldsymbol{x} = \alpha \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix},
\end{aligned}$$

which yields the same result as the previous question, i.e.,

$$F \cap G = \text{span}[\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}] .$$

2.16 Are the following mappings linear?
Recall: To show that Φ is a linear mapping from E to F, we need to show that for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in E and all λ in $\mathbb{R}$:

- $\Phi(\boldsymbol{x} + \boldsymbol{y}) = \Phi(\boldsymbol{x}) + \Phi(\boldsymbol{y})$
- $\Phi(\lambda \boldsymbol{x}) = \lambda \Phi(\boldsymbol{x})$

a. Let $a, b \in \mathbb{R}$.

$$\Phi : L^1([a, b]) \to \mathbb{R}$$
$$f \mapsto \Phi(f) = \int_a^b f(x)dx ,$$

where $L^1([a, b])$ denotes the set of integrable functions on $[a, b]$.

- Let $f, g \in L^1([a, b])$. It holds that

$$\Phi(f)+\Phi(g) = \int_a^b f(x)dx + \int_a^b g(x)dx = \int_a^b f(x)+g(x)dx = \Phi(f+g) .$$

- For $\lambda \in \mathbb{R}$ we have

$$\Phi(\lambda f) = \int_a^b \lambda f(x)dx = \lambda \int_a^b f(x)dx = \lambda \Phi(f) .$$

Therefore, Φ is a linear mapping. (In more advanced courses/books, you may learn that Φ is a linear *functional*, i.e., it takes functions as arguments. But for our purposes here this is not relevant.)

b.

$$\Phi : C^1 \to C^0$$
$$f \mapsto \Phi(f) = f' ,$$

where for $k \geqslant 1$, C^k denotes the set of k times continuously differentiable functions, and C^0 denotes the set of continuous functions.

- For $f, g \in C^1$ we have

$$\Phi(f + g) = (f + g)' = f' + g' = \Phi(f) + \Phi(g)$$

- For $\lambda \in \mathbb{R}$ we have

$$\Phi(\lambda f) = (\lambda f)' = \lambda f' = \lambda \Phi(f)$$

Therefore, Φ is linear. (Again, Φ is a linear functional.)
From the first two exercises, we have seen that both integration and differentiation are linear operations.

c.

$$\begin{aligned}\Phi : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \Phi(x) = \cos(x)\end{aligned}$$

We have $\cos(\pi) = -1$ and $\cos(2\pi) = 1$m which is different from $2\cos(\pi)$. Therefore, Φ is not linear.

d.

$$\begin{aligned}\Phi : \mathbb{R}^3 &\to \mathbb{R}^2 \\ \boldsymbol{x} &\mapsto \begin{bmatrix}1 & 2 & 3\\1 & 4 & 3\end{bmatrix}\boldsymbol{x}\end{aligned}$$

We define the matrix as $\boldsymbol{A}$. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be in $\mathbb{R}^3$. Let λ be in $\mathbb{R}$. Then:

$$\Phi(\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{A}(\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}\boldsymbol{y} = \Phi(\boldsymbol{x}) + \Phi(\boldsymbol{y})\,.$$

Similarly,

$$\Phi(\lambda\boldsymbol{x}) = \boldsymbol{A}(\lambda\boldsymbol{x}) = \lambda\boldsymbol{A}\boldsymbol{x} = \lambda\Phi(\boldsymbol{x})\,.$$

Therefore, this mapping is linear.

e. Let θ be in $[0, 2\pi[$ and

$$\begin{aligned}\Phi : \mathbb{R}^2 &\to \mathbb{R}^2 \\ \boldsymbol{x} &\mapsto \begin{bmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{bmatrix}\boldsymbol{x}\end{aligned}$$

We define the (rotation) matrix as $\boldsymbol{A}$. Then the reasoning is identical to the previous question. Therefore, this mapping is linear.
The mapping Φ represents a *rotation* of $\boldsymbol{x}$ by an angle θ. Rotations are also linear mappings.

2.17 Consider the linear mapping

$$\begin{aligned}\Phi : \mathbb{R}^3 &\to \mathbb{R}^4 \\ \Phi\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right) &= \begin{bmatrix}3x_1 + 2x_2 + x_3\\ x_1 + x_2 + x_3\\ x_1 - 3x_2\\ 2x_1 + 3x_2 + x_3\end{bmatrix}\end{aligned}$$

- Find the transformation matrix $\boldsymbol{A}_\Phi$.
- Determine $\mathrm{rk}(\boldsymbol{A}_\Phi)$.
- Compute the kernel and image of Φ. What are $\dim(\ker(\Phi))$ and $\dim(\mathrm{Im}(\Phi))$?

- The transformation matrix is

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

- The rank of $\boldsymbol{A}_\Phi$ is the number linearly independent rows/columns. We use Gaussian elimination on $\boldsymbol{A}_\Phi$ to determine the reduced row echelon form (Not necessary to identify the number of linearly independent rows/columns, but useful for the next questions.):

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

  From here, we see that $\mathrm{rk}(\boldsymbol{A}_\Phi) = 3$.
- $\ker(\Phi) = \boldsymbol{0}$ and $\dim(\ker(\Phi)) = 0$. From the reduced row echelon form, we see that all three columns of $\boldsymbol{A}_\Phi$ are linearly independent. Therefore, they form a basis of $\mathrm{Im}(\Phi)$, and $\dim(\mathrm{Im}(\Phi)) = 3$.

2.18 Let E be a vector space. Let f and g be two automorphisms on E such that $f \circ g = \mathrm{id}_E$ (i.e., $f \circ g$ is the identity mapping id_E). Show that $\ker(f) = \ker(g \circ f)$, $\mathrm{Im}(g) = \mathrm{Im}(g \circ f)$ and that $\ker(f) \cap \mathrm{Im}(g) = \{\boldsymbol{0}_E\}$.
Let $\boldsymbol{x} \in \ker(f)$. We have $g(f(\boldsymbol{x})) = g(\boldsymbol{0}) = \boldsymbol{0}$ since g is linear. Therefore, $\ker(f) \subseteq \ker(g \circ f)$ (this always holds).
Let $\boldsymbol{x} \in \ker(g \circ f)$. We have $g(f(\boldsymbol{x})) = \boldsymbol{0}$ and as f is linear, $f(g(f(\boldsymbol{x}))) = f(\boldsymbol{0}) = \boldsymbol{0}$. This implies that $(f \circ g)(f(\boldsymbol{x})) = \boldsymbol{0}$ so that $f(\boldsymbol{x}) = \boldsymbol{0}$ since $f \circ g = \mathrm{id}_E$. So $\ker(g \circ f) \subseteq \ker(f)$ and thus $\ker(g \circ f) = \ker(f)$.

Let $\boldsymbol{y} \in \mathrm{Im}(g \circ f)$ and let $\boldsymbol{x} \in E$ so that $\boldsymbol{y} = (g \circ f)(\boldsymbol{x})$. Then $\boldsymbol{y} = g(f(\boldsymbol{x}))$, which shows that $\mathrm{Im}(g \circ f) \subseteq \mathrm{Im}(g)$ (which is always true).
Let $\boldsymbol{y} \in \mathrm{Im}(g)$. Let then $\boldsymbol{x} \in E$ such that $\boldsymbol{y} = g(\boldsymbol{x})$. We have $\boldsymbol{y} = g((f \circ g)(\boldsymbol{x}))$ and thus $\boldsymbol{y} = (g \circ f)(g(\boldsymbol{x}))$, which means that $\boldsymbol{y} \in \mathrm{Im}(g \circ f)$. Therefore, $\mathrm{Im}(g) \subseteq \mathrm{Im}(g \circ f)$. Overall, $\mathrm{Im}(g) = \mathrm{Im}(g \circ f)$.

Let $\boldsymbol{y} \in \ker(f) \cap \mathrm{Im}(g)$. Let $\boldsymbol{x} \in E$ such that $\boldsymbol{y} = g(\boldsymbol{x})$. Applying f gives us $f(\boldsymbol{y}) = (f \circ g)(\boldsymbol{x})$ and as $\boldsymbol{y} \in \ker(f)$, we have $\boldsymbol{0} = \boldsymbol{x}$. This means that $\ker(f) \cap \mathrm{Im}(g) \subseteq \{\boldsymbol{0}\}$ but the intersection of two subspaces is a subspace and thus always contains $\boldsymbol{0}$, so $\ker(f) \cap \mathrm{Im}(g) = \{\boldsymbol{0}\}$.

2.19 Consider an endomorphism $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ whose transformation matrix (with respect to the standard basis in $\mathbb{R}^3$) is

$$\boldsymbol{A}_\Phi = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

a. Determine $\ker(\Phi)$ and $\mathrm{Im}(\Phi)$.
The image $\mathrm{Im}(\Phi)$ is spanned by the columns of $\boldsymbol{A}$. One way to determine a basis, we need to determine the smallest generating set of the columns of $\boldsymbol{A}_\Phi$. This can be done by Gaussian elimination. However, in this case, it is quite obvious that $\boldsymbol{A}_\Phi$ has full rank, i.e., the set of columns is already minimal, such that

$$\mathrm{Im}(\Phi) = [\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}] = \mathbb{R}^3$$

We know that $\dim(\mathrm{Im}(\Phi)) = 3$. Using the rank-nullity theorem, we get that $\dim(\ker(\Phi)) = 3 - \dim(\mathrm{Im}(\Phi)) = 0$, and $\ker(\Phi) = \{\mathbf{0}\}$ consists of the $\mathbf{0}$-vector alone.

b. Determine the transformation matrix $\tilde{\boldsymbol{A}}_\Phi$ with respect to the basis

$$B = (\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}),$$

i.e., perform a basis change toward the new basis B.

Let $\boldsymbol{B}$ the matrix built out of the basis vectors of B (order is important):

$$\boldsymbol{B} = \begin{bmatrix}1 & 1 & 1\\1 & 2 & 0\\1 & 1 & 0\end{bmatrix}.$$

Then, $\tilde{\boldsymbol{A}}_\Phi = \boldsymbol{B}^{-1}\boldsymbol{A}_\Phi\boldsymbol{B}$. The inverse is given by

$$\boldsymbol{B}^{-1} = \begin{bmatrix}0 & -1 & 2\\0 & 1 & -1\\1 & 0 & -1\end{bmatrix},$$

and the desired transformation matrix of Φ with respect to the new basis B of $\mathbb{R}^3$ is

$$\begin{bmatrix}0 & -1 & 2\\0 & 1 & -1\\1 & 0 & -1\end{bmatrix}\begin{bmatrix}1 & 1 & 0\\1 & -1 & 0\\1 & 1 & 1\end{bmatrix}\begin{bmatrix}1 & 1 & 1\\1 & 2 & 0\\1 & 1 & 0\end{bmatrix} = \begin{bmatrix}1 & 3 & 2\\0 & -2 & -1\\0 & 0 & -1\end{bmatrix}\begin{bmatrix}1 & 1 & 1\\1 & 2 & 0\\1 & 1 & 0\end{bmatrix}$$
$$= \begin{bmatrix}6 & 9 & 1\\-3 & -5 & 0\\-1 & -1 & 0\end{bmatrix}.$$

2.20 Let us consider $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}'_1, \boldsymbol{b}'_2$, 4 vectors of $\mathbb{R}^2$ expressed in the standard basis of $\mathbb{R}^2$ as

$$\boldsymbol{b}_1 = \begin{bmatrix}2\\1\end{bmatrix}, \quad \boldsymbol{b}_2 = \begin{bmatrix}-1\\-1\end{bmatrix}, \quad \boldsymbol{b}'_1 = \begin{bmatrix}2\\-2\end{bmatrix}, \quad \boldsymbol{b}'_2 = \begin{bmatrix}1\\1\end{bmatrix}$$

and let us define two ordered bases $B = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ and $B' = (\boldsymbol{b}'_1, \boldsymbol{b}'_2)$ of $\mathbb{R}^2$.

a. Show that B and B' are two bases of $\mathbb{R}^2$ and draw those basis vectors.

The vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are clearly linearly independent and so are $\boldsymbol{b}'_1$ and $\boldsymbol{b}'_2$.

b. Compute the matrix $\boldsymbol{P}_1$ that performs a basis change from B' to B.

We need to express the vector $\boldsymbol{b}'_1$ (and $\boldsymbol{b}'_2$) in terms of the vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$. In other words, we want to find the real coefficients λ_1 and λ_2 such that $\boldsymbol{b}'_1 = \lambda_1\boldsymbol{b}_1 + \lambda_2\boldsymbol{b}_2$. In order to do that, we will solve the linear equation system

$$\left[\begin{array}{cc|c}\boldsymbol{b}_1 & \boldsymbol{b}_2 & \boldsymbol{b}'_1\end{array}\right]$$

i.e.,

$$\left[\begin{array}{cc|c}2 & -1 & 2\\1 & -1 & -2\end{array}\right]$$

and which results in the reduced row echelon form

$$\left[\begin{array}{cc|c} 1 & 0 & 4 \\ 0 & 1 & 6 \end{array}\right].$$

This gives us $\boldsymbol{b}_1' = 4\boldsymbol{b}_1 + 6\boldsymbol{b}_2$.
Similarly for $\boldsymbol{b}_2'$, Gaussian elimination gives us $\boldsymbol{b}_2' = -1\boldsymbol{b}_2$.
Thus, the matrix that performs a basis change from B' to B is given as

$$\boldsymbol{P}_1 = \begin{bmatrix} 4 & 0 \\ 6 & -1 \end{bmatrix}.$$

c. We consider $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$, three vectors of $\mathbb{R}^3$ defined in the standard basis of $\mathbb{R}$ as

$$\boldsymbol{c}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \boldsymbol{c}_2 = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}, \quad \boldsymbol{c}_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

and we define $C = (\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3)$.

(i) Show that C is a basis of $\mathbb{R}^3$, e.g., by using determinants (see Section 4.1).
We have:

$$\det(\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3) = \begin{vmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ -1 & 2 & -1 \end{vmatrix} = 4 \neq 0$$

Therefore, C is regular, and the columns of C are linearly independent, i.e., they form a basis of $\mathbb{R}^3$.

(ii) Let us call $C' = (\boldsymbol{c}_1', \boldsymbol{c}_2', \boldsymbol{c}_3')$ the standard basis of $\mathbb{R}^3$. Determine the matrix $\boldsymbol{P}_2$ that performs the basis change from C to C'.
In order to write the matrix that performs a basis change from C to C', we need to express the vectors of C in terms of those of C'. But as C' is the standard basis, it is straightforward that $\boldsymbol{c}_1 = 1\boldsymbol{c}_1' + 2\boldsymbol{c}_2' - 1\boldsymbol{c}_3'$ for example. Therefore,

$$\boldsymbol{P}_2 := \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ -1 & 2 & -1 \end{bmatrix}.$$

simply contains the column vectors of C (this would not be the case if C' was not the standard basis).

d. We consider a homomorphism $\Phi : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$, such that

$$\begin{aligned} \Phi(\boldsymbol{b}_1 + \boldsymbol{b}_2) &= \boldsymbol{c}_2 + \boldsymbol{c}_3 \\ \Phi(\boldsymbol{b}_1 - \boldsymbol{b}_2) &= 2\boldsymbol{c}_1 - \boldsymbol{c}_2 + 3\boldsymbol{c}_3 \end{aligned}$$

where $B = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ and $C = (\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3)$ are ordered bases of $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively.
Determine the transformation matrix $\boldsymbol{A}_\Phi$ of Φ with respect to the ordered bases B and C.

Adding and subtracting both equations gives us

$$\begin{cases} \Phi(\boldsymbol{b}_1+\boldsymbol{b}_2)+\Phi(\boldsymbol{b}_1-\boldsymbol{b}_2) &= 2\boldsymbol{c}_1+4\boldsymbol{c}_3 \\ \Phi(\boldsymbol{b}_1+\boldsymbol{b}_2)-\Phi(\boldsymbol{b}_1-\boldsymbol{b}_2) &= -2\boldsymbol{c}_1+2\boldsymbol{c}_2-2\boldsymbol{c}_3 \end{cases}$$

As Φ is linear, we obtain

$$\begin{cases} \Phi(2\boldsymbol{b}_1) &= 2\boldsymbol{c}_1+4\boldsymbol{c}_3 \\ \Phi(2\boldsymbol{b}_2) &= -2\boldsymbol{c}_1+2\boldsymbol{c}_2-2\boldsymbol{c}_3 \end{cases}$$

And by linearity of Φ again, the system of equations gives us

$$\begin{cases} \Phi(\boldsymbol{b}_1) &= \boldsymbol{c}_1+2\boldsymbol{c}_3 \\ \Phi(\boldsymbol{b}_2) &= -\boldsymbol{c}_1+\boldsymbol{c}_2-\boldsymbol{c}_3 \end{cases}.$$

Therefore, the transformation matrix of $\boldsymbol{A}_\Phi$ with respect to the bases B and C is

$$\boldsymbol{A}_\Phi = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 2 & -1 \end{bmatrix}.$$

e. Determine $\boldsymbol{A}'$, the transformation matrix of Φ with respect to the bases $\boldsymbol{B}'$ and $\boldsymbol{C}'$.
We have:

$$\boldsymbol{A}' = \boldsymbol{P}_2\boldsymbol{A}\boldsymbol{P}_1 = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 6 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ -10 & 3 \\ 12 & -4 \end{bmatrix}.$$

f. Let us consider the vector $\boldsymbol{x} \in \mathbb{R}^2$ whose coordinates in B' are $[2,3]^\top$. In other words, $\boldsymbol{x} = 2\boldsymbol{b}_1' + 3\boldsymbol{b}_2'$.

(i) Calculate the coordinates of $\boldsymbol{x}$ in B.
By definition of $\boldsymbol{P}_1$, $\boldsymbol{x}$ can be written in B as

$$\boldsymbol{P}_1 \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 6 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 8 \\ 9 \end{bmatrix}$$

(ii) Based on that, compute the coordinates of $\Phi(\boldsymbol{x})$ expressed in C.
Using the transformation matrix $\boldsymbol{A}$ of Φ with respect to the bases B and C, we get the coordinates of $\Phi(\boldsymbol{x})$ in C with

$$\boldsymbol{A} \begin{bmatrix} 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 8 \\ 9 \end{bmatrix} = \begin{bmatrix} -1 \\ 9 \\ 7 \end{bmatrix}$$

(iii) Then, write $\Phi(\boldsymbol{x})$ in terms of $\boldsymbol{c}_1', \boldsymbol{c}_2', \boldsymbol{c}_3'$.
Going back to the basis C' thanks to the matrix $\boldsymbol{P}_2$ gives us the expression of $\Phi(\boldsymbol{x})$ in C'

$$\boldsymbol{P}_2 \begin{bmatrix} -1 \\ 9 \\ 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 9 \\ 7 \end{bmatrix} = \begin{bmatrix} 6 \\ -11 \\ 12 \end{bmatrix}$$

In other words, $\Phi(\boldsymbol{x}) = 6\boldsymbol{c}_1' - 11\boldsymbol{c}_2' + 12\boldsymbol{c}_3'$.

(iv) Use the representation of $\boldsymbol{x}$ in B' and the matrix $\boldsymbol{A}'$ to find this result directly.

We can calculate $\Phi(\boldsymbol{x})$ in C directly:

$$\boldsymbol{A}' \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ -10 & 3 \\ 12 & -4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ -11 \\ 12 \end{bmatrix}$$

3

Analytic Geometry

Exercises

3.1 Show that $\langle \cdot, \cdot \rangle$ defined for all $\boldsymbol{x} = [x_1, x_2]^\top \in \mathbb{R}^2$ and $\boldsymbol{y} = [y_1, y_2]^\top \in \mathbb{R}^2$ by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2(x_2 y_2)$$

is an inner product.

We need to show that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is a symmetric, positive definite bilinear form.

- Let $\boldsymbol{x} := [x_1, x_2]^\top$, $\boldsymbol{y} = [y_1, y_2]^\top \in \mathbb{R}^2$. Then,

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{y} \rangle &= x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2x_2 y_2 \\ &= y_1 x_1 - (y_1 x_2 + y_2 x_1) + 2 y_2 x_2 = \langle \boldsymbol{y}, \boldsymbol{x} \rangle \ , \end{aligned}$$

  where we exploited the commutativity of addition and multiplication in $\mathbb{R}$. Therefore, $\langle \cdot, \cdot \rangle$ is symmetric.
- It holds that

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = x_1^2 - (2x_1 x_2) + 2x_2^2 = (x_1 - x_2)^2 + x_2^2 \,.$$

  This is a sum of positive terms for $\boldsymbol{x} \neq \boldsymbol{0}$. Moreover, this expression shows that if $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ then $x_2 = 0$ and then $x_1 = 0$, i.e., $\boldsymbol{x} = \boldsymbol{0}$. Hence, $\langle \cdot, \cdot \rangle$ is positive definite.
- In order to show that $\langle \cdot, \cdot \rangle$ is bilinear (linear in both arguments), we will simply show that $\langle \cdot, \cdot \rangle$ is linear in its first argument. Symmetry will ensure that $\langle \cdot, \cdot \rangle$ is bilinear. Do not duplicate the proof of linearity in both arguments.

  Let $\boldsymbol{z} = [z_1, z_2]^\top \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$. Then,

$$\begin{aligned} \langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle &= (x_1 + y_1) z_1 - \big((x_1 + y_1) z_2 + (x_2 + y_2) z_2 \big) + 2\big(x_2 + y_2 \big) z_2 \big) \\ &= x_1 z_1 - (x_1 z_2 + x_2 z_1) + 2(x_2 z_2) + y_1 z_1 - (y_1 z_2 + y_2 z_1) \\ &\quad + 2(y_2 z_2) \\ &= \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle \\ \langle \lambda \boldsymbol{x}, \boldsymbol{y} \rangle &= \lambda x_1 y_1 - (\lambda x_1 y_2 + \lambda x_2 y_1) + 2(\lambda x_2 y_2) \\ &= \lambda \big(x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2(x_2 y_2) \big) = \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle \end{aligned}$$

  Thus, $\langle \cdot, \cdot \rangle$ is linear in its first variable. By symmetry, it is bilinear. Overall, $\langle \cdot, \cdot \rangle$ is an inner product.

3.2 Consider $\mathbb{R}^2$ with $\langle \cdot, \cdot \rangle$ defined for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^2$ as

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \underbrace{\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}}_{=:\boldsymbol{A}} \boldsymbol{y} .$$

Is $\langle \cdot, \cdot \rangle$ an inner product?
Let us define $\boldsymbol{x}$ and $\boldsymbol{y}$ as

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

We have $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$ but $\langle \boldsymbol{y}, \boldsymbol{x} \rangle = 1$, i.e., $\langle \cdot, \cdot \rangle$ is not symmetric. Therefore, it is not an inner product.

3.3 Compute the distance between

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

using

a. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \boldsymbol{y}$

b. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{y}\,, \quad \boldsymbol{A} := \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

The difference vector is

$$\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} .$$

a. $\|\boldsymbol{z}\| = \sqrt{\boldsymbol{z}^\top \boldsymbol{z}} = \sqrt{29}$

b. $\|\boldsymbol{z}\| = \sqrt{\boldsymbol{z}^\top \boldsymbol{A} \boldsymbol{z}} = \sqrt{55}$

3.4 Compute the angle between

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

using

a. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \boldsymbol{y}$

b. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \boldsymbol{B} \boldsymbol{y}\,, \quad \boldsymbol{B} := \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$

It holds that

$$\cos \omega = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \, \|\boldsymbol{y}\|} ,$$

where ω is the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$.

a.

$$\cos \omega = \frac{-3}{\sqrt{5}\sqrt{2}} = -\frac{3}{\sqrt{10}} \approx 2.82 \,\mathrm{rad} = 161.5^\circ$$

b.

$$\cos \omega = \frac{\boldsymbol{x}^\top \boldsymbol{B} \boldsymbol{y}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{B} \boldsymbol{y}}\sqrt{\boldsymbol{y}^\top \boldsymbol{B} \boldsymbol{y}}} = \frac{-11}{\sqrt{18}\sqrt{7}} = -\frac{11}{126} \approx 1.66 \,\mathrm{rad} = 95^\circ$$

3.5 Consider the Euclidean vector space $\mathbb{R}^5$ with the dot product. A subspace $U \subseteq \mathbb{R}^5$ and $\boldsymbol{x} \in \mathbb{R}^5$ are given by

$$U = \text{span}[\begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -3 \\ 5 \\ 0 \\ 7 \end{bmatrix}], \quad \boldsymbol{x} = \begin{bmatrix} -1 \\ -9 \\ -1 \\ 4 \\ 1 \end{bmatrix}$$

a. Determine the orthogonal projection $\pi_U(\boldsymbol{x})$ of $\boldsymbol{x}$ onto U
First, we determine a basis of U. Writing the spanning vectors as the columns of a matrix $\boldsymbol{A}$, we use Gaussian elimination to bring $\boldsymbol{A}$ into (reduced) row echelon form:

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From here, we see that the first three columns are pivot columns, i.e., the first three vectors in the generating set of U form a basis of U:

$$U = \text{span}[\begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \\ 2 \\ 1 \end{bmatrix}].$$

Now, we define

$$\boldsymbol{B} = \begin{bmatrix} 0 & 1 & -3 \\ -1 & -3 & 4 \\ 2 & 1 & 1 \\ 0 & -1 & 2 \\ 2 & 2 & 1 \end{bmatrix},$$

where we define three basis vectors $\boldsymbol{b}_i$ of U as the columns of $\boldsymbol{B}$ for $1 \leqslant i \leqslant 3$.
We know that the projection of $\boldsymbol{x}$ on U exists and we define $\boldsymbol{p} := \pi_U(\boldsymbol{x})$. Moreover, we know that $\boldsymbol{p} \in U$. We define $\boldsymbol{\lambda} := [\lambda_1, \lambda_2, \lambda_3]^\top \in \mathbb{R}^3$, such that $\boldsymbol{p}$ can be written $\boldsymbol{p} = \sum_{i=1}^{3} \lambda_i \boldsymbol{b}_i = \boldsymbol{B\lambda}$.
As $\boldsymbol{p}$ is the orthogonal projection of $\boldsymbol{x}$ onto U, then $\boldsymbol{x} - \boldsymbol{p}$ is orthogonal to all the basis vectors of U, so that

$$\boldsymbol{B}^\top(\boldsymbol{x} - \boldsymbol{B\lambda}) = \boldsymbol{0}.$$

Therefore,

$$\boldsymbol{B}^\top \boldsymbol{B\lambda} = \boldsymbol{B}^\top \boldsymbol{x}.$$

Solving in $\boldsymbol{\lambda}$ the inhomogeneous system $\boldsymbol{B}^\top \boldsymbol{B\lambda} = \boldsymbol{B}^\top \boldsymbol{x}$ gives us a single solution

$$\boldsymbol{\lambda} = \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}$$

and, therefore, the desired projection

$$\boldsymbol{p} = \boldsymbol{B}\boldsymbol{\lambda} = \begin{bmatrix} 1 \\ -5 \\ -1 \\ -2 \\ 3 \end{bmatrix} \in U\,.$$

b. Determine the distance $d(\boldsymbol{x}, U)$
The distance is simply the length of $\boldsymbol{x} - \boldsymbol{p}$:

$$\|\boldsymbol{x} - \boldsymbol{p}\| = \left\| \begin{bmatrix} 2 \\ 4 \\ 0 \\ -6 \\ 2 \end{bmatrix} \right\| = \sqrt{60}$$

3.6 Consider $\mathbb{R}^3$ with the inner product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^\top \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \boldsymbol{y}\,.$$

Furthermore, we define $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ as the standard/canonical basis in $\mathbb{R}^3$.

a. Determine the orthogonal projection $\pi_U(\boldsymbol{e}_2)$ of $\boldsymbol{e}_2$ onto

$$U = \operatorname{span}[\boldsymbol{e}_1, \boldsymbol{e}_3]\,.$$

Hint: Orthogonality is defined through the inner product.
Let $\boldsymbol{p} = \pi_U(\boldsymbol{e}_2)$. As $\boldsymbol{p} \in U$, we can define $\Lambda = (\lambda_1, \lambda_3) \in \mathbb{R}^2$ such that $\boldsymbol{p}$ can be written $\boldsymbol{p} = U\Lambda$. In fact, $\boldsymbol{p}$ becomes $\boldsymbol{p} = \lambda_1\boldsymbol{e}_1 + \lambda_3\boldsymbol{e}_3 = [\lambda_1, 0, \lambda_3]^\top$ expressed in the canonical basis.
Now, we know by orthogonal projection that

$$\begin{aligned} \boldsymbol{p} = \pi_U(\boldsymbol{e}_2) &\implies (\boldsymbol{p} - \boldsymbol{e}_2) \perp U \\ &\implies \begin{bmatrix} \langle \boldsymbol{p} - \boldsymbol{e}_2, \boldsymbol{e}_1 \rangle \\ \langle \boldsymbol{p} - \boldsymbol{e}_2, \boldsymbol{e}_3 \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &\implies \begin{bmatrix} \langle \boldsymbol{p}, \boldsymbol{e}_1 \rangle - \langle \boldsymbol{e}_2, \boldsymbol{e}_1 \rangle \\ \langle \boldsymbol{p}, \boldsymbol{e}_3 \rangle - \langle \boldsymbol{e}_2, \boldsymbol{e}_3 \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned}$$

We compute the individual components as

$$\langle \boldsymbol{p}, \boldsymbol{e}_1 \rangle = \begin{bmatrix} \lambda_1 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 2\lambda_1$$

$$\langle \boldsymbol{p}, \boldsymbol{e}_3 \rangle = \begin{bmatrix} \lambda_1 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 2\lambda_3$$

$$\langle \boldsymbol{e}_2, \boldsymbol{e}_1 \rangle = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1$$

$$\langle \boldsymbol{e}_2, \boldsymbol{e}_3 \rangle = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = -1$$

This now leads to the inhomogeneous linear equation system

$$2\lambda_1 = 1$$
$$2\lambda_3 = -1$$

This immediately gives the coordinates of the projection as

$$\pi_U(\boldsymbol{e}_2) = \begin{bmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{bmatrix}$$

b. Compute the distance $d(\boldsymbol{e}_2, U)$.
The distance of $d(\boldsymbol{e}_2, U)$ is the distance between $\boldsymbol{e}_2$ and its orthogonal projection $\boldsymbol{p} = \pi_U(\boldsymbol{e}_2)$ onto U. Therefore,

$$d(\boldsymbol{e}_2, U) = \sqrt{\langle \boldsymbol{p} - \boldsymbol{e}_2, \boldsymbol{p} - \boldsymbol{e}_2 \rangle^2}\,.$$

However,

$$\langle \boldsymbol{p} - \boldsymbol{e}_2, \boldsymbol{p} - \boldsymbol{e}_2 \rangle = \begin{bmatrix} \frac{1}{2} & -1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ -1 \\ -\frac{1}{2} \end{bmatrix} = 1\,,$$

which yields $d(\boldsymbol{e}_2, U) = \sqrt{\langle \boldsymbol{p} - \boldsymbol{e}_2, \boldsymbol{p} - \boldsymbol{e}_2 \rangle} = 1$

c. Draw the scenario: standard basis vectors and $\pi_U(\boldsymbol{e}_2)$
See Figure 3.1.

3.7 Let V be a vector space and π an endomorphism of V.

a. Prove that π is a projection if and only if $\mathrm{id}_V - \pi$ is a projection, where id_V is the identity endomorphism on V.
b. Assume now that π is a projection. Calculate $\mathrm{Im}(\mathrm{id}_V - \pi)$ and $\ker(\mathrm{id}_V - \pi)$ as a function of $\mathrm{Im}(\pi)$ and $\ker(\pi)$.

a. It holds that $(\mathrm{id}_V - \pi)^2 = \mathrm{id}_V - 2\pi + \pi^2$. Therefore,

$$(\mathrm{id}_V - \pi)^2 = \mathrm{id}_V - \pi \iff \pi^2 = \pi\,,$$

which is exactly what we want. Note that we reasoned directly at the endomorphism level, but one can also take any $\boldsymbol{x} \in V$ and prove the same results. Also note that π^2 means $\pi \circ \pi$ as in "π composed with π".

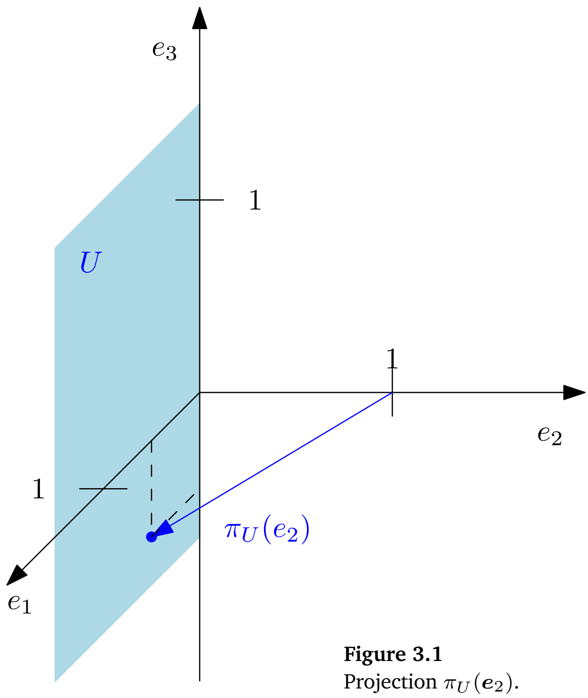


Figure 3.1
Projection $\pi_U(\boldsymbol{e}_2)$.

b. We have $\pi \circ (\mathrm{id}_V - \pi) = \pi - \pi^2 = 0_V$, where 0_V represents the null endomorphism. Then $\mathrm{Im}(\mathrm{id}_V - \pi) \subseteq \ker(\pi)$.
Conversely, let $\boldsymbol{x} \in \ker(\pi)$. Then

$$(\mathrm{id}_V - \pi)(\boldsymbol{x}) = \boldsymbol{x} - \pi(\boldsymbol{x}) = \boldsymbol{x}\,,$$

which means that $\boldsymbol{x}$ is the image of itself by $\mathrm{id}_V - \pi$. Hence, $\boldsymbol{x} \in \mathrm{Im}(\mathrm{id}_V - \pi)$. In other words, $\ker(\pi) \subseteq \mathrm{Im}(\mathrm{id}_V - \pi)$ and thus $\ker(\pi) = \mathrm{Im}(\mathrm{id}_V - \pi)$.

Similarly, we have

$$(\mathrm{id}_V - \pi) \circ \pi = \pi - \pi^2 = \pi - \pi = 0_V$$

so $\mathrm{Im}(\pi) \subseteq \ker(\mathrm{id}_V - \pi)$.
Conversely, let $\boldsymbol{x} \in \ker(\mathrm{id}_V - \pi)$. We have $(\mathrm{id}_V - \pi)(\boldsymbol{x}) = \boldsymbol{0}$ and thus $\boldsymbol{x} - \pi(\boldsymbol{x}) = \boldsymbol{0}$ or $\boldsymbol{x} = \pi(\boldsymbol{x})$. This means that $\boldsymbol{x}$ is its own image by π, and therefore $\ker(\mathrm{id}_V - \pi) \subseteq \mathrm{Im}(\pi)$. Overall,

$$\ker(\mathrm{id}_V - \pi) = \mathrm{Im}(\pi)\,.$$

3.8 Using the Gram-Schmidt method, turn the basis $B = (\boldsymbol{b}_1, \boldsymbol{b}_2)$ of a two-dimensional subspace $U \subseteq \mathbb{R}^3$ into an ONB $C = (\boldsymbol{c}_1, \boldsymbol{c}_2)$ of U, where

$$\boldsymbol{b}_1 := \begin{bmatrix}1\\1\\1\end{bmatrix}, \quad \boldsymbol{b}_2 := \begin{bmatrix}-1\\2\\0\end{bmatrix}.$$

We start by normalizing $\boldsymbol{b}_1$

$$\boldsymbol{c}_1 := \frac{\boldsymbol{b}_1}{\|\boldsymbol{b}_1\|} = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}. \tag{3.1}$$

To get $\boldsymbol{c}_2$, we project $\boldsymbol{b}_2$ onto the subspace spanned by $\boldsymbol{c}_1$. This gives us (since $\|\boldsymbol{c}_1 = 1\|$)

$$\boldsymbol{c}_1^\top \boldsymbol{b}_2 \boldsymbol{c}_1 = \frac{1}{3}\begin{bmatrix}1\\1\\1\end{bmatrix} \in U.$$

By subtracting this projection (a multiple of $\boldsymbol{c}_1$) from $\boldsymbol{b}_2$, we get a vector that is orthogonal to $\boldsymbol{c}_1$:

$$\begin{bmatrix}-1\\2\\0\end{bmatrix} - \frac{1}{3}\begin{bmatrix}1\\1\\1\end{bmatrix} = \frac{1}{3}\begin{bmatrix}-4\\5\\-1\end{bmatrix} = \frac{1}{3}(-\boldsymbol{b}_1 + 3\boldsymbol{b}_2) \in U.$$

Normalizing $\tilde{\boldsymbol{c}}_2$ yields

$$\boldsymbol{c}_2 = \frac{\tilde{\boldsymbol{c}}_2}{\|\tilde{\boldsymbol{c}}_2\|} = \frac{3}{\sqrt{42}}\begin{bmatrix}-4\\5\\-1\end{bmatrix}.$$

We see that $\boldsymbol{c}_1 \perp \boldsymbol{c}_2$ and that $\|\boldsymbol{c}_1\| = 1 = \|\boldsymbol{c}_2\|$. Moreover, $\boldsymbol{c}_1, \boldsymbol{c}_2 \in U$ it follows that $(\boldsymbol{c}_1, \boldsymbol{c}_2)$ are a basis of U.

3.9 Let $n \in \mathbb{N}$ and let $x_1, \ldots, x_n > 0$ be n positive real numbers so that $x_1 + \cdots + x_n = 1$. Use the Cauchy-Schwarz inequality and show that

a. $\sum_{i=1}^n x_i^2 \geqslant \frac{1}{n}$
b. $\sum_{i=1}^n \frac{1}{x_i} \geqslant n^2$

Hint: Think about the dot product on $\mathbb{R}^n$. Then, choose specific vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and apply the Cauchy-Schwarz inequality.

Recall Cauchy-Schwarz inequality expressed with the dot product in $\mathbb{R}^n$. Let $\boldsymbol{x} = [x_1, \ldots, x_n]^\top$ and $\boldsymbol{y} = [y_1, \ldots, y_n]^\top$ be two vectors of $\mathbb{R}^n$. Cauchy-Schwarz tells us that

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle^2 \leqslant \langle \boldsymbol{x}, \boldsymbol{x}\rangle \cdot \langle \boldsymbol{y}, \boldsymbol{y}\rangle,$$

which, applied with the dot product in $\mathbb{R}^n$, can be rephrased as

$$\left(\sum_{i=1}^n x_i y_i\right)^2 \leqslant \left(\sum_{i=1}^n x_i^2\right) \cdot \left(\sum_{i=1}^n y_i^2\right).$$

a. Consider $\boldsymbol{x} = [x_1, \ldots, x_n]^\top$ as defined in the question. Let us choose $\boldsymbol{y} = [1, \ldots, 1]^\top$. Then, the Cauchy-Schwarz inequality becomes

$$\left(\sum_{i=1}^n x_i \cdot 1\right)^2 \leqslant \left(\sum_{i=1}^n x_i^2\right) \cdot \left(\sum_{i=1}^n 1^2\right)$$

and thus

$$1 \leqslant \left(\sum_{i=1}^n x_i^2\right) \cdot n,$$

which yields the expected result.

b. Let us now choose both vectors differently to obtain the expected result. Let $\boldsymbol{x} = [\frac{1}{\sqrt{x_1}}, \dots, \frac{1}{\sqrt{x_n}}]^\top$ and $\boldsymbol{y} = [\sqrt{x_1}, \cdots, \sqrt{n}]^\top$. Note that our choice is legal since all x_i and y_i are strictly positive. The Cauchy-Schwarz inequality now becomes

$$\left(\sum_{i=1}^{n} \frac{1}{\sqrt{x_i}} \cdot \sqrt{x_i}\right)^2 \leqslant \left(\sum_{i=1}^{n} \left(\frac{1}{\sqrt{x_i}}\right)^2\right) \cdot \left(\sum_{i=1}^{n} \sqrt{x_i}^2\right)$$

so that

$$n^2 \leqslant \left(\sum_{i=1}^{n} \frac{1}{x_i}\right) \cdot \left(\sum_{i=1}^{n} x_i\right) .$$

This yields $n^2 \leqslant \sum_{i=1}^{n} \frac{1}{x_i} \cdot 1$, which gives the expected result.

3.10 Rotate the vectors

$$\boldsymbol{x}_1 := \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \boldsymbol{x}_2 := \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

by 30°.

Since $30^\circ = \pi/6\,\mathrm{rad}$ we obtain the rotation matrix

$$\boldsymbol{A} = \begin{bmatrix} \cos(\frac{\pi}{6}) & -\sin(\frac{\pi}{6}) \\ \sin(\frac{\pi}{6}) & \cos(\frac{\pi}{6}) \end{bmatrix}$$

and the rotated vectors are

$$\boldsymbol{A}\boldsymbol{x}_1 = \begin{bmatrix} 2\cos(\frac{\pi}{6}) - 3\sin(\frac{\pi}{6}) \\ 2\sin(\frac{\pi}{6}) + 3\cos(\frac{\pi}{6}) \end{bmatrix} \approx \begin{bmatrix} 0.23 \\ 3.60 \end{bmatrix}$$

$$\boldsymbol{A}\boldsymbol{x}_2 = \begin{bmatrix} \sin(\frac{\pi}{6}) \\ -\cos(\frac{\pi}{6}) \end{bmatrix} \approx \begin{bmatrix} 0.5 \\ 0.87 \end{bmatrix} .$$

4

Matrix Decompositions

Exercises

4.1 Compute the determinant using the Laplace expansion (using the first row) and the Sarrus rule for

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \\ 0 & 2 & 4 \end{bmatrix}.$$

The determinant is

$$\begin{aligned} |A| = \det(\boldsymbol{A}) &= \begin{vmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \\ 0 & 2 & 4 \end{vmatrix} \\ &= 1\begin{vmatrix} 4 & 6 \\ 2 & 4 \end{vmatrix} - 3\begin{vmatrix} 2 & 6 \\ 0 & 4 \end{vmatrix} + 5\begin{vmatrix} 2 & 4 \\ 0 & 2 \end{vmatrix} \\ &= 1 \cdot 4 - 3 \cdot 8 - 5 \cdot 4 = 0 \quad \text{Laplace expansion} \\ &= 16 + 20 + 0 - 0 - 12 - 24 = 0 \quad \text{Sarrus' rule} \end{aligned}$$

4.2 Compute the following determinant efficiently:

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

This strategy shows the power of the methods we learned in this and the previous chapter. We can first apply Gaussian elimination to transform $\boldsymbol{A}$ into a triangular form, and then use the fact that the determinant of a triangular matrix equals the product of its diagonal elements.

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

$$= \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & -1 & 4 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 & -3 \end{vmatrix} = 6\,.$$

Alternatively, we can apply the Laplace expansion and arrive at the same solution:

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

$$\overset{\text{1st col.}}{=} (-1)^{1+1} 2 \cdot \begin{vmatrix} -1 & -1 & -1 & 1 \\ 1 & 2 & 1 & 2 \\ 0 & 3 & 1 & 2 \\ 0 & -1 & -1 & 1 \end{vmatrix}\,.$$

If we now subtract the fourth row from the first row and multiply (-2) times the third column to the fourth column we obtain

$$2 \begin{vmatrix} -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & -1 & -1 & 3 \end{vmatrix} \overset{\text{1st row}}{=} -2 \begin{vmatrix} 2 & 1 & 0 \\ 3 & 1 & 0 \\ -1 & -1 & 3 \end{vmatrix} \overset{\text{3rd col.}}{=} (-2) \cdot 3(-1)^{3+3} \cdot \begin{vmatrix} 2 & 1 \\ 3 & 1 \end{vmatrix} = 6\,.$$

4.3 Compute the eigenspaces of

a.

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

b.

$$\boldsymbol{B} := \begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix}$$

a. For

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

(i) Characteristic polynomial: $p(\lambda) = |\boldsymbol{A} - \lambda \boldsymbol{I}_2| = \begin{vmatrix} 1-\lambda & 0 \\ 1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2$. Therefore $\lambda = 1$ is the only root of p and, therefore, the only eigenvalue of $\boldsymbol{A}$

(ii) To compute the eigenspace for the eigenvalue $\lambda = 1$, we need to compute the null space of $\boldsymbol{A} - \boldsymbol{I}$:

$$(\boldsymbol{A} - 1 \cdot \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \iff \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}$$

$$\Rightarrow E_1 = [\begin{bmatrix} 0 \\ 1 \end{bmatrix}]$$

E_1 is the only eigenspace of $\boldsymbol{A}$.

b. For $\boldsymbol{B}$, the corresponding eigenspaces (for $\lambda_i \in \mathbb{C}$) are

$$E_i = [\begin{bmatrix} 1 \\ i \end{bmatrix}], \quad E_{-i} = [\begin{bmatrix} 1 \\ -i \end{bmatrix}].$$

4.4 Compute all eigenspaces of

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

(i) Characteristic polynomial:

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

$$= \begin{vmatrix} -\lambda & -1-\lambda & 0 & 1 \\ 0 & -\lambda & -1-\lambda & 3-\lambda \\ 0 & 1 & -1-\lambda & 2\lambda \\ 1 & 0 & 0 & -\lambda \end{vmatrix}$$

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

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

$$= (1+\lambda)^2(\lambda^2 - 3\lambda + 2) = (1+\lambda)^2(1-\lambda)(2-\lambda)$$

Therefore, the eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = -1, \lambda_2 = 1, \lambda_3 = 2$.

(ii) The corresponding eigenspaces are the solutions of $(\boldsymbol{A} - \lambda_i \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$, $i = 1, 2, 3$, and given by

$$E_{-1} = \text{span}[\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}], \quad E_1 = \text{span}[\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}], \quad E_2 = \text{span}[\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}].$$

4.5 Diagonalizability of a matrix is unrelated to its invertibility. Determine for the following four matrices whether they are diagonalizable and/or invertible

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

In the four matrices above, the first one is diagonalizable and invertible, the second one is diagonalizable but is not invertible, the third one is invertible but is not diagonalizable, and, finally, the fourth ione s neither invertible nor is it diagonalizable.

4.6 Compute the eigenspaces of the following transformation matrices. Are they diagonalizable?

a. For

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 & 0 \\ 1 & 4 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

(i) Compute the eigenvalue as the roots of the characteristic polynomial

$$p(\lambda) = \det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \begin{vmatrix} 2-\lambda & 3 & 0 \\ 1 & 4-\lambda & 3 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)\begin{vmatrix} 2-\lambda & 3 \\ 1 & 4-\lambda \end{vmatrix}$$
$$= (1-\lambda)\big((2-\lambda)(4-\lambda)-3\big) = (1-\lambda)(8-2\lambda-4\lambda+\lambda^2-3)$$
$$= (1-\lambda)(\lambda^2-6\lambda+5) = (1-\lambda)(\lambda-1)(\lambda-5) = -(1-\lambda)^2(\lambda-5)\,.$$

Therefore, we obtain the eigenvalues 1 and 5.

(ii) To compute the eigenspaces, we need to solve $(\boldsymbol{A}-\lambda_i\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$, where $\lambda_1 = 1, \lambda_2 = 5$:

$$E_1: \begin{bmatrix} 1 & 3 & 0 \\ 1 & 3 & 3 \\ 0 & 0 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

where we subtracted the first row from the second and, subsequently, divided the second row by 3 to obtain the reduced row echelon form. From here, we see that

$$E_1 = \mathrm{span}[\begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}]$$

Now, we compute E_5 by solving $(\boldsymbol{A}-5\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$:

$$\begin{bmatrix} -3 & 3 & 0 \\ 1 & -1 & 3 \\ 0 & 0 & -4 \end{bmatrix} \begin{array}{l} |\cdot(-\frac{1}{3}) \\ +\frac{1}{3}R_1 + \frac{3}{4}R_3 \\ |\cdot(-\frac{1}{4}), \text{swap with } R_2 \end{array} \rightsquigarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Then,

$$E_5 = \mathrm{span}[\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}]\,.$$

(iii) This endomorphism cannot be diagonalized because $\dim(E_1) + \dim(E_5) \neq 3$.

Alternative arguments:

- $\dim(E_1)$ does not correspond to the algebraic multiplicity of the eigenvalue $\lambda = 1$ in the characteristic polynomial
- $\mathrm{rk}(\boldsymbol{A}-\boldsymbol{I}) \neq 3-2$.

b. For

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

(i) Compute the eigenvalues as the roots of the characteristic polynomial

$$p(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = (1-\lambda)\lambda^3$$
$$= \begin{vmatrix} 1-\lambda & 1 & 0 & 0 \\ 0 & -\lambda & 0 & 0 \\ 0 & 0 & -\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{vmatrix} = -(1-\lambda)\lambda^3 .$$

It follows that the eigenvalues are 0 and 1 with algebraic multiplicities 3 and 1, respectively.

(ii) We compute the eigenspaces E_0 and E_1, which requires us to determine the null spaces $\boldsymbol{A} - \lambda_i \boldsymbol{I}$, where $\lambda_i \in \{0, 1\}$.

(iii) We compute the eigenspaces E_0 and E_1, which requires us to determine the null spaces $\boldsymbol{A} - \lambda_i \boldsymbol{I}$, where $\lambda_i \in \{0, 1\}$. For E_0, we compute the null space of $\boldsymbol{A}$ directly and obtain

$$E_0 = \text{span}[\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}] .$$

To determine E_1, we need to solve $(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{array}{l} \\ +R_1 | \text{ move to } R_4 \\ \cdot(-1) \\ \cdot(-1) \end{array} \rightsquigarrow \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From here, we see that

$$E_1 = \text{span}[\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}] .$$

(iv) Since $\dim(E_0) + \dim(E_1) = 4 = \dim(\mathbb{R}^4)$, it follows that a diagonal form exists.

4.7 Are the following matrices diagonalizable? If yes, determine their diagonal form and a basis with respect to which the transformation matrices are diagonal. If no, give reasons why they are not diagonalizable.

a.

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix}$$

We determine the characteristic polynomial as

$$p(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = -\lambda(4-\lambda) + 8 = \lambda^2 - 4\lambda + 8\,.$$

The characteristic polynomial does not decompose into linear factors over $\mathbb{R}$ because the roots of $p(\lambda)$ are complex and given by $\lambda_{1,2} = 2 \pm \sqrt{-4}$. Since the characteristic polynomial does not decompose into linear factors, $\boldsymbol{A}$ cannot be diagonalized (over $\mathbb{R}$).

b.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

(i) The characteristic polynomial is $p(\lambda) = \det(\boldsymbol{A} - \lambda I)$.

$$p(\lambda) = \begin{vmatrix} 1-\lambda & 1 & 1 \\ 1 & 1-\lambda & 1 \\ 1 & 1 & 1-\lambda \end{vmatrix} \overset{\text{subtr. } R_1 \text{ from } R_2, R_3}{=} \begin{vmatrix} 1-\lambda & 1 & 1 \\ \lambda & -\lambda & 0 \\ \lambda & 0 & -\lambda \end{vmatrix}$$
$$\overset{\text{develop last row}}{=} \lambda \begin{vmatrix} 1 & 1 \\ -\lambda & 0 \end{vmatrix} - \lambda \begin{vmatrix} 1-\lambda & 1 \\ \lambda & -\lambda \end{vmatrix}$$
$$= \lambda^2 + \lambda(\lambda(1-\lambda) + \lambda) = \lambda(-\lambda^2 + 3\lambda) = \lambda^2(\lambda - 3)\,.$$

Therefore, the roots of $p(\lambda)$ are 0 and 3 with algebraic multiplicities 2 and 1, respectively.

(ii) To determine whether $\boldsymbol{A}$ is diagonalizable, we need to show that the dimension of E_0 is 2 (because the dimension of E_3 is necessarily 1: an eigenspace has at least dimension 1 by definition, and its dimension cannot exceed the algebraic multiplicity of its associated eigenvalue).

Let us study $E_0 = \ker(\boldsymbol{A} - 0\boldsymbol{I})$:

$$\boldsymbol{A} - 0\boldsymbol{I} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Here, $\dim E_0 = 2$, which is identical to the algebraic multiplicity of the eigenvalue 0 in the characteristic polynomial. Thus A is diagonalizable. Moreover, we can read from the reduced row echelon form that

$$E_0 = \operatorname{span}[\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}].$$

(iii) For E_3, we obtain

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

which has rank 2, and, therefore (using the rank-nullity theorem),

E_3 has dimension 1 (it could not be anything else anyway, as justified above) and

$$E_1 = \text{span}[\begin{bmatrix}-1\\-1\\-1\end{bmatrix}].$$

(iv) Therefore, we can find a new basis P as the concatenation of the spanning vectors of the eigenspaces. If we identify the matrix

$$\boldsymbol{P} = \begin{bmatrix}-1 & 1 & 1\\-1 & -1 & 0\\-1 & 0 & -1\end{bmatrix},$$

whose columns are composed of the basis vectors of basis P, then our endomorphism will have the diagonal form: $\boldsymbol{D} = \text{diag}[3,0,0]$ with respect to this new basis. As a reminder, $\text{diag}[3,0,0]$ refers to the 3×3 diagonal matrix with $3,0,0$ as values on the diagonal.
Note that the diagonal form is not unique and depends on the order of the eigenvectors in the new basis. For example, we can define another matrix $\boldsymbol{Q}$ composed of the same vectors as $\boldsymbol{P}$ but in a different order:

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

If we use this matrix, our endomorphism would have another diagonal form: $\boldsymbol{D}' = \text{diag}[0,3,0]$. Sceptical students can check that $\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{D}'$ and $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}$.

c.

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

$p(\lambda) = (\lambda-1)(\lambda-2)(\lambda-4)^2$

$$E_1 = \text{span}[\begin{bmatrix}-1\\1\\0\\0\end{bmatrix}], \quad E_2 = \text{span}[\begin{bmatrix}1\\-1\\0\\1\end{bmatrix}], \quad E_4 = \text{span}[\begin{bmatrix}1\\0\\-1\\1\end{bmatrix}].$$

Here, we see that $\dim(E_4) = 1 \neq 2$ (which is the algebraic multiplicity of the eigenvalue 4). Therefore, $\boldsymbol{A}$ cannot be diagonalized.

d.

$$\boldsymbol{A} = \begin{bmatrix}5 & -6 & -6\\-1 & 4 & 2\\3 & -6 & -4\end{bmatrix}$$

(i) We compute the characteristic polynomial $p(\lambda) = \det(\boldsymbol{A} - \lambda\boldsymbol{I})$ as

$$\begin{aligned} p(\lambda) &= \begin{vmatrix} 5-\lambda & -6 & -6 \\ -1 & 4-\lambda & 2 \\ 3 & -6 & -4-\lambda \end{vmatrix} \\ &= (5-\lambda)(4-\lambda)(-4-\lambda) - 36 - 36 + 18(4-\lambda) \\ &\quad + 12(5-\lambda) - 6(-4-\lambda) \\ &= -\lambda^3 + 5\lambda^2 - 8\lambda + 4 = (1-\lambda)(2-\lambda)^2 , \end{aligned}$$

where we used Sarrus rule. The characteristic polynomial decomposes into linear factors, and the eigenvalues are $\lambda_1 = 1, \lambda_2 = 2$ with (algebraic) multiplicity 1 and 2, respectively.

(ii) If the dimension of the eigenspaces are identical to multiplicity of the corresponding eigenvalues, the matrix is diagonalizable. The eigenspace dimension is the dimension of $\ker(\boldsymbol{A} - \lambda_i\boldsymbol{I})$, where λ_i are the eigenvalues (here: 1,2). For a simple check whether the matrices are diagonalizable, it is sufficient to compute the rank r_i of $\boldsymbol{A} - \lambda_i\boldsymbol{I}$ since the eigenspace dimension is $n - r_i$ (rank-nullity theorem).
Let us study E_2 and apply Gaussian elimination on $\boldsymbol{A} - 2\boldsymbol{I}$:

$$\begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix} . \tag{4.1}$$

(iii) We can immediately see that the rank of this matrix is 1 since the first and third row are three times the second. Therefore, the eigenspace dimension is $\dim(E_2) = 3 - 1 = 2$, which corresponds to the algebraic multiplicity of the eigenvalue $\lambda = 2$ in $p(\lambda)$. Moreover, we know that the dimension of E_1 is 1 since it cannot exceed its algebraic multiplicity, and the dimension of an eigenspace is at least 1. Hence, $\boldsymbol{A}$ is diagonalizable.

(iv) The diagonal matrix is easy to determine since it just contains the eigenvalues (with corresponding multiplicities) on its diagonal:

$$\boldsymbol{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} .$$

(v) We need to determine a basis with respect to which the transformation matrix is diagonal. We know that the basis that consists of the eigenvectors has exactly this property. Therefore, we need to determine the eigenvectors for all eigenvalues. Remember that $\boldsymbol{x}$ is an eigenvector for an eigenvalue λ if they satisfy $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} \iff (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$. Therefore, we need to find the basis vectors of the eigenspaces E_1, E_2.
For $E_1 = \ker(\boldsymbol{A} - \boldsymbol{I})$ we obtain:

$$\begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix} \begin{matrix} +4R_2 \\ \\ +3R_2 \end{matrix} \rightsquigarrow \begin{bmatrix} 0 & 6 & 2 \\ -1 & 3 & 2 \\ 0 & 3 & 1 \end{bmatrix} \begin{matrix} \cdot(\frac{1}{6}) \\ \cdot(-1)|\text{swap with } R_1 \\ -\frac{1}{2}R_1 \end{matrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & -3 & -2 \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} +3R_2 \\ \\ \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix}$$

The rank of this matrix is 2. Since $3 - 2 = 1$ it follows that $\dim(E_1) = 1$, which corresponds to the algebraic multiplicity of the eigenvalue $\lambda = 1$ in the characteristic polynomial.

(vi) From the reduced row echelon form we see that

$$E_1 = \text{span}[\begin{bmatrix} 3 \\ -1 \\ 3 \end{bmatrix}],$$

and our first eigenvector is $[3, -1, 3]^\top$.

(vii) We proceed with determining a basis of E_2, which will give us the other two basis vectors that we need (remember that $\dim(E_2) = 2$). From (4.1), we (almost) immediately obtain the reduced row echelon form

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

and the corresponding eigenspace

$$E_2 = \text{span}[\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}].$$

(viii) Overall, an ordered basis with respect to which $\boldsymbol{A}$ has diagonal form $\boldsymbol{D}$ consists of all eigenvectors is

$$B = (\begin{bmatrix} 3 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}).$$

4.8 Find the SVD of the matrix

$$\boldsymbol{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}.$$

$$\boldsymbol{A} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{=\boldsymbol{U}} \underbrace{\begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}}_{=\boldsymbol{\Sigma}} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} & -\frac{2}{3} \\ \frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} & \frac{2}{3} \\ 0 & -\frac{2\sqrt{2}}{3} & \frac{1}{3} \end{bmatrix}}_{=\boldsymbol{V}}$$

4.9 Find the singular value decomposition of

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

(i) Compute the symmetrized matrix $A^\top A$. We first compute the symmetrized matrix

$$A^\top A = \begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} . \tag{4.2}$$

(ii) Find the right-singular vectors and singular values from $A^\top A$.
The characteristic polynomial of $A^\top A$ is

$$(5-\lambda)^2 - 9 = \lambda^2 - 10\lambda + 16 = 0 . \tag{4.3}$$

This yields eigenvalues, sorted from largest absolute first, $\lambda_1 = 8$ and $\lambda_2 = 2$. The associated normalized eigenvectors are respectively

$$v_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad \text{and} \quad v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} . \tag{4.4}$$

We have thus obtained the right-singular orthogonal matrixx

$$V = [v_1, v_2] = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} . \tag{4.5}$$

(iii) Determine the singular values.
We obtain the two singular values from the square root of the eigenvalues $\sigma_1 = \sqrt{\lambda_1} = \sqrt{8} = 2\sqrt{2}$ and $\sigma_2 = \sqrt{\lambda_2} = \sqrt{2}$. We construct the singular value diagonal matrix as

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} . \tag{4.6}$$

(iv) Find the left-singular eigenvectors.
We have to map the two eigenvectors v_1, v_2 using A. This yields two self-consistent equations that enable us to find orthogonal eigenvectors u_1, u_2:

$$A v_1 = (\sigma_1 u_1 v_1^\top) v_1 = \sigma_1 u_1 (v_1^\top v_1) = \sigma_1 u_1 = \begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix}$$

$$A v_2 = (\sigma_2 u_2 v_2^\top) v_2 = \sigma_2 u_2 (v_2^\top v_2) = \sigma_2 u_2 = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}$$

We normalize the left-singular vectors by diving them by their respective singular values and obtain $u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, which yields

$$U = [u_1, u_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} . \tag{4.7}$$

(v) Assemble the left-/right-singular vectors and singular values. The SVD of A is

$$A = U \Sigma V^\top = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{8} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} .$$

4.10 Find the rank-1 approximation of

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

To find the rank-1 approximation we apply the SVD to $\boldsymbol{A}$ (as in Exercise 4.7) to obtain

$$\boldsymbol{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} & -\frac{2}{3} \\ \frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} & \frac{2}{3} \\ 0 & -\frac{2\sqrt{2}}{3} & \frac{1}{3} \end{bmatrix}.$$

We apply the construction rule for rank-1 matrices

$$\boldsymbol{A}_i = \boldsymbol{u}_i \boldsymbol{v}_i^\top .$$

We use the largest singular value ($\sigma_1 = 5$, i.e., $i = 1$ and the first column vectors of the $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices, respectively:

$$\boldsymbol{A}_1 = \boldsymbol{u}_1 \boldsymbol{v}_1^\top = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

To find the rank-1 approximation, we apply the SVD to $\boldsymbol{A}$ to obtain

$$\boldsymbol{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\boldsymbol{V} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} & -\frac{2}{3} \\ \frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} & \frac{2}{3} \\ 0 & -\frac{2\sqrt{2}}{3} & \frac{1}{3} \end{bmatrix}.$$

We apply the construction rule for rank-1 matrices

$$\boldsymbol{A}_i = \boldsymbol{u}_i \boldsymbol{v}_i^\top .$$

We use the largest singular value ($\sigma_1 = 5$, i.e., $i = 1$) and therefore, the first column vectors of $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively, which then yields

$$\boldsymbol{A}_1 = \boldsymbol{u}_1 \boldsymbol{v}_1^\top = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

4.11 Show that for any $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ the matrices $\boldsymbol{A}^\top \boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^\top$ possess the same nonzero eigenvalues.

Let us assume that λ is a nonzero eigenvalue of $\boldsymbol{A}\boldsymbol{A}^\top$ and $\boldsymbol{x}$ is an eigenvector belonging to λ. Thus, the eigenvalue equation

$$(\boldsymbol{A}\boldsymbol{A}^\top)\boldsymbol{x} = \lambda \boldsymbol{x}$$

can be manipulated by left multiplying by $\boldsymbol{A}^\top$ and pulling on the right-hand side the scalar factor λ forward. This yields

$$\boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top)\boldsymbol{x} = \boldsymbol{A}^\top (\lambda \boldsymbol{x}) = \lambda (\boldsymbol{A}^\top \boldsymbol{x})$$

and we can use matrix multiplication associativity to reorder the left-hand side factors

$$(\boldsymbol{A}^\top \boldsymbol{A})(\boldsymbol{A}^\top \boldsymbol{x}) = \lambda(\boldsymbol{A}^\top \boldsymbol{x})\,.$$

This is the eigenvalue equation for $\boldsymbol{A}^\top \boldsymbol{A}$. Therefore, λ is the same eigenvalue for $\boldsymbol{A}\boldsymbol{A}^\top$ and $\boldsymbol{A}^\top \boldsymbol{A}$.

4.12 Show that for $\boldsymbol{x} \neq \boldsymbol{0}$ Theorem 4.24 holds, i.e., show that

$$\max_{\boldsymbol{x}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \sigma_1\,,$$

where σ_1 is the largest singular value of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$.

(i) We compute the eigendecomposition of the symmetric matrix

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top$$

for diagonal $\boldsymbol{D}$ and orthogonal $\boldsymbol{P}$. Since the columns of $\boldsymbol{P}$ are an ONB of $\mathbb{R}^n$, we can write every $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{x}$ as a linear combination of the eigenvectors $\boldsymbol{p}_i$ so that

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{p}_i\,, \quad \boldsymbol{x} \in \mathbb{R}\,. \tag{4.8}$$

Moreover, since the orthogonal matrix $\boldsymbol{P}$ preserves lengths (see Section 3.4), we obtain

$$\|\boldsymbol{y}\|_2^2 = \|\boldsymbol{P}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2 = \sum_{i=1}^{n} x_i^2\,. \tag{4.9}$$

(ii) Then,

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = \boldsymbol{x}^\top(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top)\boldsymbol{x} = \boldsymbol{y}^\top \boldsymbol{D}\boldsymbol{y} = \left\langle \sum_{i=1}^{n} \sqrt{\lambda_i} x_i \boldsymbol{p}_i, \sum_{i=1}^{n} \sqrt{\lambda_i} x_i \boldsymbol{p}_i \right\rangle\,,$$

where we used $\langle\cdot,\cdot\rangle$ to denote the dot product.

(iii) The bilinearity of the dot product gives us

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = \sum_{i=1}^{n} \lambda_i \langle x_i \boldsymbol{p}_i, x_i \boldsymbol{p}_i\rangle = \sum_{i=1}^{n} \lambda_i x_i^2$$

where we exploited that the $\boldsymbol{p}_i$ are an ONB and $\boldsymbol{p}_i^\top \boldsymbol{p}_i = 1$.

(iv) With (4.8) we obtain

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leqslant \big(\max_{1\leqslant j\leqslant n} \lambda_j\big) \sum_{i=1}^{n} x_i^2 \overset{(4.9)}{=} \max_{1\leqslant j\leqslant n} \lambda_j \|\boldsymbol{x}\|_2^2$$

so that

$$\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \leqslant \max_{1\leqslant j\leqslant n} \lambda_j\,,$$

where λ_j are the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$.

(v) Assuming the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$ are sorted in descending order, we get

$$\frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leqslant \sqrt{\lambda_1} = \sigma_1 \,,$$

where σ_1 is the maximum singular value of $\boldsymbol{A}$.

5

Vector Calculus

Exercises

5.1 Compute the derivative $f'(x)$ for

$$f(x) = \log(x^4)\sin(x^3).$$

$$f'(x) = \frac{4}{x}\sin(x^3) + 12x^2\log(x)\cos(x^3)$$

5.2 Compute the derivative $f'(x)$ of the logistic sigmoid

$$f(x) = \frac{1}{1+\exp(-x)}.$$

$$f'(x) = \frac{\exp(x)}{(1+\exp(x))^2}$$

5.3 Compute the derivative $f'(x)$ of the function

$$f(x) = \exp(-\tfrac{1}{2\sigma^2}(x-\mu)^2),$$

where $\mu,\ \sigma \in \mathbb{R}$ are constants.

$$f'(x) = -\tfrac{1}{\sigma^2}f(x)(x-\mu)$$

5.4 Compute the Taylor polynomials T_n, $n = 0, \ldots, 5$ of $f(x) = \sin(x) + \cos(x)$ at $x_0 = 0$.

$$\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= T_0(x) + x \\
T_2(x) &= T_1(x) - \frac{x^2}{2} \\
T_3(x) &= T_2(x) - \frac{x^3}{6} \\
T_4(x) &= T_3(x) + \frac{x^4}{24} \\
T_5(x) &= T_4(x) + \frac{x^5}{120}
\end{aligned}$$

5.5 Consider the following functions:

$$\begin{aligned} f_1(\boldsymbol{x}) &= \sin(x_1)\cos(x_2)\,, \quad \boldsymbol{x} \in \mathbb{R}^2 \\ f_2(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{x}^\top \boldsymbol{y}\,, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \\ f_3(\boldsymbol{x}) &= \boldsymbol{x}\boldsymbol{x}^\top\,, \quad \boldsymbol{x} \in \mathbb{R}^n \end{aligned}$$

a. What are the dimensions of $\frac{\partial f_i}{\partial \boldsymbol{x}}$?
b. Compute the Jacobians.

- f_1

$$\begin{aligned} \frac{\partial f_1}{\partial x_1} &= \cos(x_1)\cos(x_2) \\ \frac{\partial f_1}{\partial x_2} &= -\sin(x_1)\sin(x_2) \\ \implies J &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \cos(x_1)\cos(x_2) & -\sin(x_1)\sin(x_2) \end{bmatrix} \in \mathbb{R}^{1\times 2} \end{aligned}$$

- f_2

$$\boldsymbol{x}^\top \boldsymbol{y} = \sum_i x_i y_i$$

$$\frac{\partial f_2}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_2}{\partial x_n} \end{bmatrix} = \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix} = \boldsymbol{y}^\top \in \mathbb{R}^n$$

$$\begin{aligned} \frac{\partial f_2}{\partial \boldsymbol{y}} &= \begin{bmatrix} \frac{\partial f_2}{\partial y_1} & \cdots & \frac{\partial f_2}{\partial y_n} \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \boldsymbol{x}^\top \in \mathbb{R}^n \\ \implies J &= \begin{bmatrix} \frac{\partial f_2}{\partial \boldsymbol{x}} & \frac{\partial f_2}{\partial \boldsymbol{y}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}^\top & \boldsymbol{x}^\top \end{bmatrix} \in \mathbb{R}^{1\times 2n} \end{aligned}$$

- $f_3 : \mathbb{R}^n \to \mathbb{R}^{n\times n}$

$$\begin{aligned} \boldsymbol{x}\boldsymbol{x}^\top &= \begin{bmatrix} x_1\boldsymbol{x}^\top \\ x_2\boldsymbol{x}^\top \\ \vdots \\ x_n\boldsymbol{x}^\top \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}x_1 & \boldsymbol{x}x_2 & \cdots & \boldsymbol{x}x_n \end{bmatrix} \in \mathbb{R}^{n\times n} \\ \implies \frac{\partial f_3}{\partial x_1} &= \underbrace{\begin{bmatrix} \boldsymbol{x}^\top \\ \boldsymbol{0}_n^\top \\ \vdots \\ \boldsymbol{0}_n^\top \end{bmatrix}}_{\in\mathbb{R}^{n\times n}} + \underbrace{\begin{bmatrix} \boldsymbol{x} & \boldsymbol{0}_n & \cdots & \boldsymbol{0}_n \end{bmatrix}}_{\in\mathbb{R}^{n\times n}} \in \mathbb{R}^{n\times n} \\ \implies \frac{\partial f_3}{\partial x_i} &= \underbrace{\begin{bmatrix} \boldsymbol{0}_{(i-1)\times n} \\ \boldsymbol{x}^\top \\ \boldsymbol{0}_{(n-1+1)\times n} \end{bmatrix}}_{\in\mathbb{R}^{n\times n}} + \underbrace{\begin{bmatrix} \boldsymbol{0}_{n\times(i-1)} & \boldsymbol{x} & \boldsymbol{0}_{n\times(n-i+1)} \end{bmatrix}}_{\in\mathbb{R}^{n\times n}} \in \mathbb{R}^{n\times n} \end{aligned}$$

To get the Jacobian, we need to concatenate all partial derivatives $\frac{\partial f_3}{\partial x_i}$ and obtain

$$J = \begin{bmatrix} \frac{\partial f_3}{\partial x_1} & \cdots & \frac{\partial f_3}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{(n\times n)\times n}$$

5.6 Differentiate f with respect to $\boldsymbol{t}$ and g with respect to $\boldsymbol{X}$, where

$$f(\boldsymbol{t}) = \sin(\log(\boldsymbol{t}^\top \boldsymbol{t}))\,, \qquad t \in \mathbb{R}^D$$
$$g(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{AXB})\,, \qquad \boldsymbol{A} \in \mathbb{R}^{D\times E}, \boldsymbol{X} \in \mathbb{R}^{E\times F}, \boldsymbol{B} \in \mathbb{R}^{F\times D}\,,$$

where tr(·) denotes the trace.

-
$$\frac{\partial f}{\partial \boldsymbol{t}} = \cos(\log(\boldsymbol{t}^\top \boldsymbol{t})) \cdot \frac{1}{\boldsymbol{t}^\top \boldsymbol{t}} \cdot 2\boldsymbol{t}^\top$$

- The trace for $\boldsymbol{T} \in \mathbb{R}^{D\times D}$ is defined as
$$\operatorname{tr}(\boldsymbol{T}) = \sum_{i=1}^{D} T_{ii}$$
A matrix product $\boldsymbol{ST}$ can be written as
$$(\boldsymbol{ST})_{pq} = \sum_i S_{pi} T_{iq}$$
The product $\boldsymbol{AXB}$ contains the elements
$$(\boldsymbol{AXB})_{pq} = \sum_{i=1}^{E} \sum_{j=1}^{F} A_{pi} X_{ij} B_{jq}$$
When we compute the trace, we sum up the diagonal elements of the matrix. Therefore we obtain,
$$\operatorname{tr}(\boldsymbol{AXB}) = \sum_{k=1}^{D} (\boldsymbol{AXB})_{kk} = \sum_{k=1}^{D} \left(\sum_{i=1}^{E} \sum_{j=1}^{F} A_{ki} X_{ij} B_{jk} \right)$$
$$\frac{\partial}{\partial X_{ij}} \operatorname{tr}(\boldsymbol{AXB}) = \sum_k A_{ki} B_{jk} = (\boldsymbol{BA})_{ji}$$
We know that the size of the gradient needs to be of the same size as $\boldsymbol{X}$ (i.e., $E \times F$). Therefore, we have to transpose the result above, such that we finally obtain
$$\frac{\partial}{\partial \boldsymbol{X}} \operatorname{tr}(\boldsymbol{AXB}) = \underbrace{\boldsymbol{A}^\top}_{E\times D} \underbrace{\boldsymbol{B}^\top}_{D\times F}$$

5.7 Compute the derivatives $\mathrm{d}f/\mathrm{d}\boldsymbol{x}$ of the following functions by using the chain rule. Provide the dimensions of every single partial derivative. Describe your steps in detail.

a.
$$f(z) = \log(1 + z)\,, \quad z = \boldsymbol{x}^\top \boldsymbol{x}\,, \quad \boldsymbol{x} \in \mathbb{R}^D$$

b.
$$f(\boldsymbol{z}) = \sin(\boldsymbol{z})\,, \quad \boldsymbol{z} = \boldsymbol{Ax} + \boldsymbol{b}\,, \quad \boldsymbol{A} \in \mathbb{R}^{E\times D}, \boldsymbol{x} \in \mathbb{R}^D, \boldsymbol{b} \in \mathbb{R}^E$$

where $\sin(\cdot)$ is applied to every element of $\boldsymbol{z}$.

a.

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \underbrace{\frac{\partial f}{\partial z}}_{\in\mathbb{R}} \underbrace{\frac{\partial z}{\partial \boldsymbol{x}}}_{\in\mathbb{R}^{1\times D}} \in \mathbb{R}^{1\times D}$$

$$\frac{\partial f}{\partial z} = \frac{1}{1+z} = \frac{1}{1+\boldsymbol{x}^\top\boldsymbol{x}}$$

$$\frac{\partial z}{\partial \boldsymbol{x}} = 2\boldsymbol{x}^\top$$

$$\implies \frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \frac{2\boldsymbol{x}^\top}{1+\boldsymbol{x}^\top\boldsymbol{x}}$$

b.

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \underbrace{\frac{\partial f}{\partial \boldsymbol{z}}}_{\in\mathbb{R}^{E\times E}} \underbrace{\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}}_{\in\mathbb{R}^{E\times D}} \in \mathbb{R}^{E\times D}$$

$$\sin(\boldsymbol{z}) = \begin{bmatrix}\sin z_1\\ \vdots \\ \sin z_E\end{bmatrix}$$

$$\frac{\partial \sin \boldsymbol{z}}{\partial z_i} = \begin{bmatrix}0\\ \vdots\\ 0\\ \cos(z_i)\\ 0\\ \vdots\\ 0\end{bmatrix} \in \mathbb{R}^E$$

$$\implies \frac{\partial f}{\partial \boldsymbol{z}} = \mathrm{diag}(\cos(\boldsymbol{z})) \in \mathbb{R}^{E\times E}$$

$$\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \boldsymbol{A} \in \mathbb{R}^{E\times D}:$$

$$c_i = \sum_{j=1}^{D} A_{ij}x_j \implies \frac{\partial c_i}{\partial x_j} = A_{ij}\,, \quad i = 1,\dots,E,\, j = 1,\dots,D$$

Here, we defined c_i to be the ith component of $\boldsymbol{A}\boldsymbol{x}$. The offset $\boldsymbol{b}$ is constant and vanishes when taking the gradient with respect to $\boldsymbol{x}$. Overall, we obtain

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \mathrm{diag}(\cos(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}))\boldsymbol{A}$$

5.8 Compute the derivatives $\mathrm{d}f/\mathrm{d}\boldsymbol{x}$ of the following functions. Describe your steps in detail.

a. Use the chain rule. Provide the dimensions of every single partial derivative.

$$f(z) = \exp(-\tfrac{1}{2}z)$$
$$z = g(\boldsymbol{y}) = \boldsymbol{y}^\top\boldsymbol{S}^{-1}\boldsymbol{y}$$
$$\boldsymbol{y} = h(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\mu}$$

where $\boldsymbol{x}, \boldsymbol{\mu} \in \mathbb{R}^D$, $\boldsymbol{S} \in \mathbb{R}^{D \times D}$.
The desired derivative can be computed using the chain rule:

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \underbrace{\frac{\partial f}{\partial z}}_{1\times 1} \underbrace{\frac{\partial g}{\partial \boldsymbol{y}}}_{1\times D} \underbrace{\frac{\partial h}{\partial \boldsymbol{x}}}_{D\times D} \in \mathbb{R}^{1\times D}$$

Here

$$\begin{aligned}
\frac{\partial f}{\partial z} &= -\tfrac{1}{2}\exp(-\tfrac{1}{2}z) \\
\frac{\partial g}{\partial \boldsymbol{y}} &= 2\boldsymbol{y}^\top \boldsymbol{S}^{-1} \\
\frac{\partial h}{\partial \boldsymbol{x}} &= \boldsymbol{I}_D
\end{aligned}$$

so that

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = -\exp\big(-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{S}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big)(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{S}^{-1}$$

b.

$$f(\boldsymbol{x}) = \mathrm{tr}(\boldsymbol{x}\boldsymbol{x}^\top + \sigma^2 \boldsymbol{I})\,, \quad \boldsymbol{x} \in \mathbb{R}^D$$

Here $\mathrm{tr}(\boldsymbol{A})$ is the trace of $\boldsymbol{A}$, i.e., the sum of the diagonal elements A_{ii}.
Hint: Explicitly write out the outer product.
Let us have a look at the outer product. We define $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top$ with

$$X_{ij} = x_i x_j$$

The trace sums up all the diagonal elements, such that

$$\frac{\partial}{\partial x_j}\mathrm{tr}(\boldsymbol{X} + \sigma^2 \boldsymbol{I}) = \sum_{i=1}^{D} \frac{\partial X_{ii} + \sigma^2}{\partial x_j} = 2x_j$$

for $j = 1, \dots, D$. Overall, we get

$$\frac{\partial}{\partial \boldsymbol{x}}\mathrm{tr}(\boldsymbol{x}\boldsymbol{x}^\top + \sigma^2 \boldsymbol{I}) = 2\boldsymbol{x}^\top \in \mathbb{R}^{1\times D}$$

c. Use the chain rule. Provide the dimensions of every single partial derivative. You do not need to compute the product of the partial derivatives explicitly.

$$\begin{aligned}
\boldsymbol{f} &= \tanh(\boldsymbol{z}) \in \mathbb{R}^M \\
\boldsymbol{z} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \quad \boldsymbol{x} \in \mathbb{R}^N, \boldsymbol{A} \in \mathbb{R}^{M\times N}, \boldsymbol{b} \in \mathbb{R}^M.
\end{aligned}$$

Here, $\tanh$ is applied to every component of $\boldsymbol{z}$.

$$\begin{aligned}
\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}} &= \mathrm{diag}(1 - \tanh^2(\boldsymbol{z})) \in \mathbb{R}^{M\times M} \\
\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} &= \frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A} \in \mathbb{R}^{M\times N}
\end{aligned}$$

We get the latter result by defining $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, such that

$$y_i = \sum_j A_{ij} x_j \implies \frac{\partial y_i}{\partial x_k} = A_{ik} \implies \frac{\partial y_i}{\partial \boldsymbol{x}} = [A_{i1}, ..., A_{iN}] \in \mathbb{R}^{1 \times N}$$
$$\implies \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \boldsymbol{A}$$

The overall derivative is an $M \times N$ matrix.

5.9 We define

$$\begin{aligned} g(\boldsymbol{z}, \boldsymbol{\nu}) &:= \log p(\boldsymbol{x}, \boldsymbol{z}) - \log q(\boldsymbol{z}, \boldsymbol{\nu}) \\ \boldsymbol{z} &:= t(\boldsymbol{\epsilon}, \boldsymbol{\nu}) \end{aligned}$$

for differentiable functions p, q, t. By using the chain rule, compute the gradient

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\nu}} g(\boldsymbol{z}, \boldsymbol{\nu}) .$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\nu}} g(\boldsymbol{z}, \boldsymbol{\nu}) &= \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\nu}} (\log p(\boldsymbol{x}, \boldsymbol{z}) - \log q(\boldsymbol{z}, \boldsymbol{\nu})) = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\nu}} \log p(\boldsymbol{x}, \boldsymbol{z}) - \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\nu}} \log q(\boldsymbol{z}, \boldsymbol{\nu}) \\ &= \frac{\partial}{\partial \boldsymbol{z}} \log p(\boldsymbol{x}, \boldsymbol{z}) \frac{\partial t(\boldsymbol{\epsilon}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} - \frac{\partial}{\partial \boldsymbol{z}} \log q(\boldsymbol{z}, \boldsymbol{\nu}) \frac{\partial t(\boldsymbol{\epsilon}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} - \frac{\partial}{\partial \boldsymbol{\nu}} \log q(\boldsymbol{z}, \boldsymbol{\nu}) \\ &= \left(\frac{\partial}{\partial \boldsymbol{z}} \log p(\boldsymbol{x}, \boldsymbol{z}) - \frac{\partial}{\partial \boldsymbol{z}} \log q(\boldsymbol{z}, \boldsymbol{\nu}) \right) \frac{\partial t(\boldsymbol{\epsilon}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} - \frac{\partial}{\partial \boldsymbol{\nu}} \log q(\boldsymbol{z}, \boldsymbol{\nu}) \\ &= \left(\frac{1}{p(\boldsymbol{x}, \boldsymbol{z})} \frac{\partial}{\partial \boldsymbol{z}} p(\boldsymbol{x}, \boldsymbol{z}) - \frac{1}{q(\boldsymbol{z}, \boldsymbol{\nu})} \frac{\partial}{\partial \boldsymbol{z}} q(\boldsymbol{z}, \boldsymbol{\nu}) \right) \frac{\partial t(\boldsymbol{\epsilon}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} - \frac{\partial}{\partial \boldsymbol{\nu}} \log q(\boldsymbol{z}, \boldsymbol{\nu}) \end{aligned}$$

6

Probability and Distributions

Exercises

6.1 Consider the following bivariate distribution $p(x, y)$ of two discrete random variables X and Y.

Y		x_1	x_2	x_3	x_4	x_5
	y_1	0.01	0.02	0.03	0.1	0.1
	y_2	0.05	0.1	0.05	0.07	0.2
	y_3	0.1	0.05	0.03	0.05	0.04

X

Compute:

a. The marginal distributions $p(x)$ and $p(y)$.
b. The conditional distributions $p(x|Y = y_1)$ and $p(y|X = x_3)$.

The marginal and conditional distributions are given by

$$
\begin{aligned}
p(x) &= [0.16, 0.17, 0.11, 0.22, 0.34]^\top \\
p(y) &= [0.26, 0.47, 0.27]^\top \\
p(x|Y = y_1) &= [0.01, 0.02, 0.03, 0.1, 0.1]^\top \\
p(y|X = x_3) &= [0.03, 0.05, 0.03]^\top .
\end{aligned}
$$

6.2 Consider a mixture of two Gaussian distributions (illustrated in Figure 6.4),

$$
0.4\,\mathcal{N}\left(\begin{bmatrix}10\\2\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + 0.6\,\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}8.4 & 2.0\\2.0 & 1.7\end{bmatrix}\right).
$$

a. Compute the marginal distributions for each dimension.
b. Compute the mean, mode and median for each marginal distribution.
c. Compute the mean and mode for the two-dimensional distribution.

Consider the mixture of two Gaussians,

$$
p\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = 0.4\,\mathcal{N}\left(\begin{bmatrix}10\\2\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + 0.6\,\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}8.4 & 2.0\\2.0 & 1.7\end{bmatrix}\right).
$$

a. Compute the marginal distribution for each dimension.

$$p(x_1) = \int 0.4 \left(\mathcal{N}\left(\begin{bmatrix}10\\2\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + 0.6\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}8.4 & 2.0\\2.0 & 1.7\end{bmatrix}\right)\right) \mathrm{d}x_2 \tag{6.1}$$

$$= 0.4 \int \mathcal{N}\left(\begin{bmatrix}10\\2\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) \mathrm{d}x_2 + 0.6 \int \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}8.4 & 2.0\\2.0 & 1.7\end{bmatrix}\right) \mathrm{d}x_2 \tag{6.2}$$

$$= 0.4\mathcal{N}\big(10,\, 1\big) + 0.6\mathcal{N}\big(0,\, 8.4\big)\,.$$

From (6.1) to (6.2), we used the result that the integral of a sum is the sum of the integrals.
Similarly, we can get

$$p(x_2) = 0.4\mathcal{N}\big(2,\, 1\big) + 0.6\mathcal{N}\big(0,\, 1.7\big)\,.$$

b. Compute the mean, mode and median for each marginal distribution.
Mean:

$$\begin{aligned}
\mathbb{E}(x_1) &= \int x_1 p(x_1) \mathrm{d}x_1 \\
&= \int x_1 \left(0.4\mathcal{N}\big(x_1 \,|\, 10,\, 1\big) + 0.6\mathcal{N}\big(x_1 \,|\, 0,\, 8.4\big)\right) \mathrm{d}x_1 && (6.3)\\
&= 0.4 \int x_1 \mathcal{N}\big(x_1 \,|\, 10,\, 1\big) \mathrm{d}x_1 + 0.6 \int x_1 \mathcal{N}\big(x_1 \,|\, 0,\, 8.4\big) \mathrm{d}x_1 && (6.4)\\
&= 0.4 \cdot 10 + 0.6 \cdot 0 = 4\,.
\end{aligned}$$

From step (6.3) to step (6.4), we use the fact that for $Y \sim \mathcal{N}\big(\mu,\, \sigma\big)$, where $\mathbb{E}[Y] = \mu$.
Similarly,

$$\mathbb{E}(x_2) = 0.4 \cdot 2 + 0.6 \cdot 0 = 0.8\,.$$

Mode: In principle, we would need to solve for

$$\frac{\mathrm{d}p(x_1)}{\mathrm{d}x_1} = 0 \quad \text{and} \quad \frac{\mathrm{d}p(x_2)}{\mathrm{d}x_2} = 0\,.$$

However, we can observe that the modes of each individual distribution are the peaks of the Gaussians, that is the Gaussian means for each dimension. Median:

$$\int_{-\infty}^{a} p(x_1)\mathrm{d}x_1 = \frac{1}{2}\,.$$

c. Compute the mean and mode for the 2 dimensional distribution.
Mean:
From (6.30), we know that

$$\mathbb{E}[\begin{bmatrix}x_1\\x_2\end{bmatrix}] = \begin{bmatrix}\mathbb{E}[x_1]\\\mathbb{E}[x_2]\end{bmatrix} = \begin{bmatrix}4\\0.8\end{bmatrix}\,.$$

Mode:
The two dimensional distribution has two peaks, and hence there are

two modes. In general, we would need to solve an optimization problem to find the maxima. However, in this particular case we can observe that the two modes correspond to the individual Gaussian means, that is

$$\begin{bmatrix}10\\2\end{bmatrix} \text{ and } \begin{bmatrix}0\\0\end{bmatrix}.$$

6.3 You have written a computer program that sometimes compiles and sometimes not (code does not change). You decide to model the apparent stochasticity (success vs. no success) x of the compiler using a Bernoulli distribution with parameter μ:

$$p(x \mid \mu) = \mu^x (1-\mu)^{1-x}, \quad x \in \{0, 1\}.$$

Choose a conjugate prior for the Bernoulli likelihood and compute the posterior distribution $p(\mu \mid x_1, \ldots, x_N)$.
The posterior is proportional to the product of the likelihood (Bernoulli) and the prior (Beta). Given the choice of the prior, we already know that the posterior will be a Beta distribution. With a conjugate Beta prior $p(\mu) = \text{Beta}(a, b)$ obtain

$$\begin{aligned} p(\mu \mid \boldsymbol{X}) &= \frac{p(\mu)p(\boldsymbol{X} \mid \mu)}{p(\boldsymbol{X})} \propto \mu^{a-1}(1-\mu)^{b-1} \prod_{i=1}^{N} \mu^{x_i}(1-\mu)^{1-x_i} \\ &\propto \mu^{a-1+\sum_i x_i}(1-\mu)^{b-1+N-\sum_i x_i} \propto \text{Beta}(a + \sum_i x_i, b + N - \sum_i x_i). \end{aligned}$$

Alternative approach: We look at the log-posterior. Ignoring all constants that are independent of μ and x_i, we obtain

$$\begin{aligned} \log p(\mu \mid \boldsymbol{X}) &= (a-1)\log\mu + (b-1)\log(1-\mu) \\ &\quad + \log\mu \sum_i x_i + \log(1-\mu)\sum_i (1-x_i) \\ &= \log\mu(a - 1 + \sum_i x_i) + \log(1-\mu)(b - 1 + N - \sum_i x_i). \end{aligned}$$

Therefore, the posterior is again $\text{Beta}(a + \sum_i x_i, b + N - \sum_i x_i)$.

6.4 There are two bags. The first bag contains four mangos and two apples; the second bag contains four mangos and four apples.
We also have a biased coin, which shows “heads” with probability 0.6 and “tails” with probability 0.4. If the coin shows “heads”. we pick a fruit at random from bag 1; otherwise we pick a fruit at random from bag 2.
Your friend flips the coin (you cannot see the result), picks a fruit at random from the corresponding bag, and presents you a mango.
What is the probability that the mango was picked from bag 2?
Hint: Use Bayes’ theorem.
We apply Bayes’ theorem and compute the posterior $p(b_2 \mid m)$ of picking a mango from bag 2.

$$p(b_2 \mid m) = \frac{p(m \mid b_2)p(b_2)}{p(m)}$$

where

$$p(m) = p(b_1)p(m \mid b_1) + p(b_2)p(m \mid b_2) = \frac{3}{5}\frac{2}{3} + \frac{2}{5}\frac{1}{2} = \frac{2}{5} + \frac{1}{5} = \frac{3}{5} \quad \text{Evidence}$$

$$p(b_2) = \tfrac{2}{5} \quad \text{Prior}$$

$$p(m \mid b_2) = \tfrac{1}{2} \quad \text{Likelihood}$$

Therefore,

$$p(b_2 \mid m) = \frac{p(m \mid b_2)p(b_2)}{p(m)} = \frac{\frac{2}{5}\frac{1}{2}}{\frac{3}{5}} = \frac{1}{3}\,.$$

6.5 Consider the time-series model

$$\begin{aligned} \boldsymbol{x}_{t+1} &= \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{w}\,, \quad \boldsymbol{w} \sim \mathcal{N}\big(\boldsymbol{0},\, \boldsymbol{Q}\big) \\ \boldsymbol{y}_t &= \boldsymbol{C}\boldsymbol{x}_t + \boldsymbol{v}\,, \quad \boldsymbol{v} \sim \mathcal{N}\big(\boldsymbol{0},\, \boldsymbol{R}\big)\,, \end{aligned}$$

where $\boldsymbol{w}, \boldsymbol{v}$ are i.i.d. Gaussian noise variables. Further, assume that $p(\boldsymbol{x}_0) = \mathcal{N}\big(\boldsymbol{\mu}_0,\, \boldsymbol{\Sigma}_0\big)$.

a. What is the form of $p(\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_T)$? Justify your answer (you do not have to explicitly compute the joint distribution).
b. Assume that $p(\boldsymbol{x}_t \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_t) = \mathcal{N}\big(\boldsymbol{\mu}_t,\, \boldsymbol{\Sigma}_t\big)$.
   1. Compute $p(\boldsymbol{x}_{t+1} \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_t)$.
   2. Compute $p(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1} \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_t)$.
   3. At time $t+1$, we observe the value $\boldsymbol{y}_{t+1} = \hat{\boldsymbol{y}}$. Compute the conditional distribution $p(\boldsymbol{x}_{t+1} \mid \boldsymbol{y}_1, \dots, \boldsymbol{y}_{t+1})$.

1. The joint distribution is Gaussian: $p(\boldsymbol{x}_0)$ is Gaussian, and $\boldsymbol{x}_{t+1}$ is a linear/affine transformations of $\boldsymbol{x}_t$. Since affine transformations leave the Gaussianity of the random variable invariant, the joint distribution must be Gaussian.
2. 1. We use the results from linear transformation of Gaussian random variables (Section 6.5.3). We immediately obtain

$$\begin{aligned} &p(\boldsymbol{x}_{t+1} \mid \boldsymbol{y}_{1:t}) = \mathcal{N}\big(\boldsymbol{x}_{t+1} \mid \boldsymbol{\mu}_{t+1\,|\,t},\, \boldsymbol{\Sigma}_{t+1\,|\,t}\big) \\ &\boldsymbol{\mu}_{t+1\,|\,t} := \boldsymbol{A}\boldsymbol{\mu}_t \\ &\boldsymbol{\Sigma}_{t+1\,|\,t} := \boldsymbol{A}\boldsymbol{\Sigma}_t\boldsymbol{A}^\top + \boldsymbol{Q}\,. \end{aligned}$$

   2. The joint distribution $p(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1} \mid \boldsymbol{y}_{1:t})$ is Gaussian (linear transformation of random variables). We compute every component of the Gaussian separately:

$$\begin{aligned} &\mathbb{E}[\boldsymbol{y}_{t+1} \mid \boldsymbol{y}_{1:t}] = \boldsymbol{C}\boldsymbol{\mu}_{t+1\,|\,t} =: \boldsymbol{\mu}^y_{t+1\,|\,t} \\ &\mathbb{V}[\boldsymbol{y}_{t+1} \mid \boldsymbol{y}_{1:t}] = \boldsymbol{C}\boldsymbol{\Sigma}_{t+1\,|\,t}\boldsymbol{C}^\top + \boldsymbol{R} =: \boldsymbol{\Sigma}^y_{t+1\,|\,t} \\ &\mathrm{Cov}[\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1} \mid \boldsymbol{y}_{1:t}] = \mathrm{Cov}[\boldsymbol{x}_{t+1}, \boldsymbol{C}\boldsymbol{x}_{t+1} \mid \boldsymbol{y}_{1:t}] = \boldsymbol{\Sigma}_{t+1\,|\,t}\boldsymbol{C}^\top =: \boldsymbol{\Sigma}_{xy} \end{aligned}$$

   Therefore,

$$p(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1} \mid \boldsymbol{y}_{1:t}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{t+1\,|\,t} \\ \boldsymbol{\mu}^y_{t+1\,|\,t} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{t+1\,|\,t} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}^y_{t+1\,|\,t} \end{bmatrix}\,.\right)$$

3. We obtain the desired distribution (again: Gaussian) by applying the rules for Gaussian conditioning of the joint distribution in B):

$$
\begin{aligned}
p(\boldsymbol{x}_{t+1} \,|\, \boldsymbol{y}_{1:t+1}) &= \mathcal{N}\big(\boldsymbol{\mu}_{t+1\,|\,t+1}, \boldsymbol{\Sigma}_{t+1\,|\,t+1}\big)\\
\boldsymbol{\mu}_{t+1\,|\,t+1} &= \boldsymbol{\mu}_{t+1\,|\,t} + \boldsymbol{\Sigma}_{xy}(\boldsymbol{\Sigma}^{y}_{t+1\,|\,t})^{-1}(\hat{\boldsymbol{y}} - \boldsymbol{\mu}^{y}_{t+1\,|\,t})\\
\boldsymbol{\Sigma}_{t+1\,|\,t+1} &= \boldsymbol{\Sigma}_{t+1\,|\,t} - \boldsymbol{\Sigma}_{xy}(\boldsymbol{\Sigma}^{y}_{t+1\,|\,t})^{-1}\boldsymbol{\Sigma}_{yx}
\end{aligned}
$$

6.6 Prove the relationship in (6.44), which relates the standard definition of the variance to the raw-score expression for the variance.
The formula in (6.43) can be converted to the so called raw-score formula for variance

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(x_i^2 - 2x_i\mu + \mu^2\right)\\
&= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \frac{2}{N}\mu\sum_{i=1}^{N}x_i + \mu^2\\
&= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \mu^2\\
&= \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N}x_i\right)^2 .
\end{aligned}
$$

6.7 Prove the relationship in (6.45), which relates the pairwise difference between examples in a dataset with the raw-score expression for the variance.
Consider the pairwise squared deviation of a random variable:

$$
\begin{aligned}
\frac{1}{N^2}\sum_{i,j=1}^{N}(x_i-x_j)^2 &= \frac{1}{N^2}\sum_{i,j=1}^{N}x_i^2 + x_j^2 - 2x_ix_j && (6.5)\\
&= \frac{1}{N}\sum_{i=1}^{N}x_i^2 + \frac{1}{N}\sum_{j=1}^{N}x_j^2 - 2\frac{1}{N^2}\sum_{i,j=1}^{N}x_ix_j && (6.6)\\
&= \frac{2}{N}\sum_{i=1}^{N}x_i^2 - 2\frac{1}{N}\sum_{i=1}^{N}x_i\left(\frac{1}{N}\sum_{j=1}^{N}x_j\right) && (6.7)\\
&= 2\left[\frac{1}{N}\sum_{i=1}^{N}x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N}x_i\right)^2\right] . && (6.8)
\end{aligned}
$$

Observe that the last equation above is twice of (6.44).
Going from (6.6) to (6.7), we need to make two observations. The first two terms in (6.6) are sums over all examples, and the index is not important, therefore they can be combined into twice the sum. The last term is obtained by "pushing in" the sum over j. This can be seen by a small example by considering $N = 3$:

$$
\begin{aligned}
&x_1x_1 + x_1x_2 + x_1x_3 + x_2x_1 + x_2x_2 + x_2x_3 + x_3x_1 + x_3x_2 + x_3x_3\\
=&x_1(x_1+x_2+x_3) + x_2(x_1+x_2+x_3) + x_3(x_1+x_2+x_3) .
\end{aligned}
$$

Going from (6.7) to (6.8), we consider the sum over j as a common factor, as seen in the following small example:

$$\begin{aligned}&x_1(x_1+x_2+x_3)+x_2(x_1+x_2+x_3)+x_3(x_1+x_2+x_3)\\=&(x_1+x_2+x_3)(x_1+x_2+x_3)\,.\end{aligned}$$

6.8 Express the Bernoulli distribution in the natural parameter form of the exponential family, see (6.107).

$$\begin{aligned}p(x\,|\,\mu)&=\mu^x(1-\mu)^{1-x}\\&=\exp\left[\log\left(\mu^x(1-\mu)^{1-x}\right)\right]\\&=\exp\left[x\log\mu+(1-x)\log(1-\mu)\right]\\&=\exp\left[x\log\frac{\mu}{1-\mu}+\log(1-\mu)\right]\\&=\exp\left[x\theta-\log(1+\exp(\theta))\right]\,,\end{aligned}$$

where the last line is obtained by substituting $\theta=\log\frac{\mu}{1-\mu}$. Note that this is in exponential family form, as θ is the natural parameter, the sufficient statistic $\phi(x)=x$ and the log partition function is $A(\theta)=\log(1+\exp(\theta))$.

6.9 Express the Binomial distribution as an exponential family distribution. Also express the Beta distribution is an exponential family distribution. Show that the product of the Beta and the Binomial distribution is also a member of the exponential family.

Express Binomial distribution as an exponential family distribution:

Recall the Binomial distribution from Example 6.11

$$p(x|N,\mu)=\binom{N}{x}\mu^x(1-\mu)^{(N-x)}\quad x=0,1,...,N\,.$$

This can be written in exponential family form

$$\begin{aligned}p(x|N,\mu)&=\binom{N}{x}\exp\left[\log(\mu^x(1-\mu)^{(N-x)})\right]\\&=\binom{N}{x}\exp\left[x\log\mu+(N-x)\log(1-\mu)\right]\\&=\binom{N}{x}\exp\left[x\log\frac{\mu}{1-\mu}+N\log(1-\mu)\right]\,.\end{aligned}$$

The last line can be identified as being in exponential family form by observing that

$$h(x)=\begin{cases}\binom{N}{x} & x=0,1,...,N.\\ 0 & \text{otherwise}\end{cases}$$

$$\theta=\log\frac{\mu}{1-\mu}$$

$$\begin{aligned}\phi(x) &= x \\ A(\theta) &= -N\log(1-\mu) = N\log(1+e^{\theta})\,.\end{aligned}$$

Express Beta distribution as an exponential family distribution:

Recall the Beta distribution from Example 6.11

$$\begin{aligned}p(\mu|\alpha,\beta) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\mu^{\alpha-1}(1-\mu)^{(\beta-1)} \\ &= \frac{1}{B(\alpha,\beta)}\mu^{\alpha-1}(1-\mu)^{(\beta-1)}\,.\end{aligned}$$

This can be written in exponential family form

$$p(\mu|\alpha,\beta) = \exp\left[(\alpha-1)\log\mu + (\beta-1)\log(1-\mu) - \log B(\alpha,\beta)\right].$$

This can be identified as being in exponential family form by observing that

$$\begin{aligned}h(\mu) &= 1 \\ \boldsymbol{\theta}|(\boldsymbol{\alpha},\boldsymbol{\beta}) &= [\alpha-1 \quad \beta-1]^\top \\ \phi(\mu) &= [\log\mu \quad \log(1-\mu)] \\ A(\theta) &= \log B(\alpha+\beta)\,.\end{aligned}$$

Show the product of the Beta and Binomial distribution is also a member of exponential family:

We treat μ as random variable here, x as an known integer between 0 and N. For $x = 0, \ldots, N$, we get

$$\begin{aligned}p(x|N,\mu)p(\mu|\alpha,\beta) &= \binom{N}{x}\mu^{x}(1-\mu)^{(N-x)}\frac{1}{B(\alpha,\beta)}\mu^{\alpha-1}(1-\mu)^{(\beta-1)} \\ &= \binom{N}{x}\exp\left[x\log\frac{\mu}{1-\mu} + N\log(1-\mu)\right] \\ &\quad\cdot\exp\left[(\alpha-1)\log\mu + (\beta-1)\log(1-\mu) - \log B(\alpha,\beta)\right] \\ &= \exp\Bigg[(x+\alpha-1)\log\mu + (N-x+\beta-1)\log(1-\mu) \\ &\qquad + \log\binom{N}{x} - \log B(\alpha,\beta)\Bigg]\,.\end{aligned}$$

The last line can be identified as being in exponential family form by observing that

$$\begin{aligned}h(\mu) &= 1 \\ \boldsymbol{\theta} &= [x+\alpha-1 \quad N-x+\beta-1]^\top \\ \phi(\mu) &= [\log\mu \quad \log(1-\mu)] \\ A(\theta) &= \log\binom{N}{x} + \log B(\alpha+\beta)\,.\end{aligned}$$

6.10 Derive the relationship in Section 6.5.2 in two ways:

a. By completing the square
b. By expressing the Gaussian in its exponential family form

The *product* of two Gaussians $\mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{a},\,\boldsymbol{A}\big)\mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{b},\,\boldsymbol{B}\big)$ is an unnormalized Gaussian distribution $c\mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{c},\,\boldsymbol{C}\big)$ with

$$
\begin{aligned}
\boldsymbol{C} &= (\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\\
\boldsymbol{c} &= \boldsymbol{C}(\boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{B}^{-1}\boldsymbol{b})\\
c &= (2\pi)^{-\frac{D}{2}}\,|\,\boldsymbol{A}+\boldsymbol{B}\,|^{-\frac{1}{2}}\exp\big(-\tfrac{1}{2}(\boldsymbol{a}-\boldsymbol{b})^\top(\boldsymbol{A}+\boldsymbol{B})^{-1}(\boldsymbol{a}-\boldsymbol{b})\big)\,.
\end{aligned}
$$

Note that the normalizing constant c itself can be considered a (normalized) Gaussian distribution either in $\boldsymbol{a}$ or in $\boldsymbol{b}$ with an "inflated" covariance matrix $\boldsymbol{A}+\boldsymbol{B}$, i.e., $c=\mathcal{N}\big(\boldsymbol{a}\,|\,\boldsymbol{b},\,\boldsymbol{A}+\boldsymbol{B}\big)=\mathcal{N}\big(\boldsymbol{b}\,|\,\boldsymbol{a},\,\boldsymbol{A}+\boldsymbol{B}\big)$.

(i) By completing the square, we obtain

$$
\begin{aligned}
\mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{a},\,\boldsymbol{A}\big)) &= (2\pi)^{-\frac{D}{2}}|\boldsymbol{A}|^{-\frac{1}{2}}\exp\,[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^\top\boldsymbol{A}^{-1}(\boldsymbol{x}-\boldsymbol{a}))],\\
\mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{b},\,\boldsymbol{B}\big)) &= (2\pi)^{-\frac{D}{2}}|\boldsymbol{B}|^{-\frac{1}{2}}\exp\,[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{b})^\top\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{b}))]\,.
\end{aligned}
$$

For convenience, we consider the log of the product:

$$
\begin{aligned}
&\log\Big(\mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{a},\,\boldsymbol{A}\big)\mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{b},\,\boldsymbol{B}\big)\Big)\\
&=\log\mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{a},\,\boldsymbol{A}\big)+\log\mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{b},\,\boldsymbol{B}\big)\\
&=-\frac{1}{2}[(\boldsymbol{x}-\boldsymbol{a})^\top\boldsymbol{A}^{-1}(\boldsymbol{x}-\boldsymbol{a})+(\boldsymbol{x}-\boldsymbol{b})^\top\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{b})]+\text{const}\\
&=-\frac{1}{2}[\boldsymbol{x}^\top\boldsymbol{A}^{-1}\boldsymbol{x}-2\boldsymbol{x}^\top\boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{a}^\top\boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{x}^\top\boldsymbol{B}^{-1}\boldsymbol{x}\\
&\qquad-2\boldsymbol{x}^\top\boldsymbol{B}^{-1}\boldsymbol{b}+\boldsymbol{b}^\top\boldsymbol{B}^{-1}\boldsymbol{b}]+\text{const}\\
&=-\frac{1}{2}[\boldsymbol{x}^\top(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}\boldsymbol{x}-2\boldsymbol{x}(\boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{B}^{-1}\boldsymbol{b})+\boldsymbol{a}^\top\boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{b}^\top\boldsymbol{B}^{-1}\boldsymbol{b}]\\
&\qquad+\text{const}\\
&=-\frac{1}{2}\Big[(\boldsymbol{x}-(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}(\boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{B}^{-1}\boldsymbol{b}))^\top\\
&\qquad\cdot(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})(\boldsymbol{x}-(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}(\boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{B}^{-1}\boldsymbol{b})\Big]+\text{const}\,.
\end{aligned}
$$

Thus, the corresponding product is $\mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{c},\,\boldsymbol{C}\big)$ with

$$\boldsymbol{C}=(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1} \tag{6.9}$$

$$\boldsymbol{c}=(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}(\boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{B}^{-1}\boldsymbol{b})=\boldsymbol{C}(\boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{B}^{-1}\boldsymbol{b})\,. \tag{6.10}$$

(ii) By expressing Gaussian in its exponential family form:

Recall from the Example 6.13 we get the exponential form of univariate Gaussian distribution, similarly we can get the exponential form of multivariate Gaussian distribution as

$$p(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{A})=\exp\Big(-\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{A}^{-1}\boldsymbol{x}+\boldsymbol{x}^\top\boldsymbol{A}^{-1}\boldsymbol{a}-\frac{1}{2}\boldsymbol{a}^\top\boldsymbol{A}^{-1}\boldsymbol{a}$$

$$- \frac{1}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{A}|\Big) .$$

This can be identified as being in exponential family form by observing that

$$\begin{aligned} h(\mu) &= 1 \\ \boldsymbol{\theta}|(\boldsymbol{a},\boldsymbol{A}) &= \begin{bmatrix} -\frac{1}{2}\mathrm{vec}\left(\boldsymbol{A}^{-1}\right) \\ \boldsymbol{A}^{-1}\boldsymbol{a} \end{bmatrix} \\ \boldsymbol{\phi}(\boldsymbol{x}) &= \begin{bmatrix} \mathrm{vec}\left(\boldsymbol{x}\boldsymbol{x}^\top\right) \\ \boldsymbol{x} \end{bmatrix} . \end{aligned}$$

Then the product of Gaussian distribution can be expressed as

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{A})p(\boldsymbol{x}|\boldsymbol{b},\boldsymbol{B}) &= \exp(\langle\boldsymbol{\theta}|(\boldsymbol{a},\boldsymbol{A}),\boldsymbol{\phi}(\boldsymbol{x})\rangle + \langle\boldsymbol{\theta}|(\boldsymbol{b},\boldsymbol{B}),\boldsymbol{\phi}(\boldsymbol{x})\rangle) + \mathrm{const} \\ &= \exp(\langle\boldsymbol{\theta}|(\boldsymbol{a},\boldsymbol{A},\boldsymbol{b},\boldsymbol{B}),\boldsymbol{\phi}(\boldsymbol{x})\rangle) + \mathrm{const}, \end{aligned}$$

where $\boldsymbol{\theta}\,|\,(\boldsymbol{a},\boldsymbol{A},\boldsymbol{b},\boldsymbol{B}) = \left[-\frac{1}{2}(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}) \quad \boldsymbol{A}^{-1}\boldsymbol{a}+\boldsymbol{B}^{-1}\boldsymbol{b}\right]^\top$.
Then we end up with the same answer as shown in (6.9)–(6.10).

6.11 **Iterated Expectations.**
Consider two random variables x, y with joint distribution $p(x, y)$. Show that

$$\mathbb{E}_X[x] = \mathbb{E}_Y\left[\mathbb{E}_X[x\,|\,y]\right] .$$

Here, $\mathbb{E}_X[x\,|\,y]$ denotes the expected value of x under the conditional distribution $p(x\,|\,y)$.

$$\begin{aligned} &\mathbb{E}_X[x] = \mathbb{E}_Y\left[\mathbb{E}_X[x\,|\,y]\right] \\ &\iff \int xp(x)dx = \iint xp(x\,|\,y)p(y)dxdy \\ &\iff \int xp(x)dx = \iint xp(x,y)dxdy \\ &\iff \int xp(x)dx = \int x\int p(x,y)dydx = \int xp(x)dx\,, \end{aligned}$$

which proves the claim.

6.12 **Manipulation of Gaussian Random Variables.**
Consider a Gaussian random variable $\boldsymbol{x} \sim \mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{\mu}_x,\,\boldsymbol{\Sigma}_x\big)$, where $\boldsymbol{x} \in \mathbb{R}^D$. Furthermore, we have

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{w}\,,$$

where $\boldsymbol{y} \in \mathbb{R}^E$, $\boldsymbol{A} \in \mathbb{R}^{E\times D}$, $\boldsymbol{b} \in \mathbb{R}^E$, and $\boldsymbol{w} \sim \mathcal{N}\big(\boldsymbol{w}\,|\,\boldsymbol{0},\,\boldsymbol{Q}\big)$ is independent Gaussian noise. "Independent" implies that $\boldsymbol{x}$ and $\boldsymbol{w}$ are independent random variables and that $\boldsymbol{Q}$ is diagonal.

a. Write down the likelihood $p(\boldsymbol{y}\,|\,\boldsymbol{x})$.

$$\begin{aligned} p(\boldsymbol{y}\,|\,\boldsymbol{x}) &= \mathcal{N}\big(\boldsymbol{y}\,|\,\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b},\,\boldsymbol{Q}\big) \\ &= Z\exp\big(-\tfrac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^\top\boldsymbol{Q}^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})\big) \\ \text{where} \quad Z &= (2\pi)^{-E/2}\,|\,\boldsymbol{Q}\,|^{-1/2} \end{aligned}$$

b. The distribution $p(\boldsymbol{y}) = \int p(\boldsymbol{y}\,|\,\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$ is Gaussian. Compute the mean $\boldsymbol{\mu}_y$ and the covariance $\boldsymbol{\Sigma}_y$. Derive your result in detail.

$p(\boldsymbol{y})$ is Gaussian distributed (affine transformation of the Gaussian random variable $\boldsymbol{x}$) with mean $\boldsymbol{\mu}_y$ and covariance matrix $\boldsymbol{\Sigma}_y$. We obtain

$$\begin{aligned}
\boldsymbol{\mu}_y &= \mathbb{E}_Y[\boldsymbol{y}] = \mathbb{E}_{X,W}[\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{w}] = \boldsymbol{A}\mathbb{E}_X[\boldsymbol{x}] + \boldsymbol{b} + \mathbb{E}_W[\boldsymbol{w}] \\
&= \boldsymbol{A}\boldsymbol{\mu}_x + \boldsymbol{b}, \quad \text{and} \\
\boldsymbol{\Sigma}_y &= \mathbb{E}_Y[\boldsymbol{y}\boldsymbol{y}^\top] - \mathbb{E}_Y[\boldsymbol{y}]\mathbb{E}_Y[\boldsymbol{y}]^\top \\
&= \mathbb{E}_{X,W}\left[(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{w})(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{w})^\top\right] - \boldsymbol{\mu}_y\boldsymbol{\mu}_y^\top \\
&= \mathbb{E}_{X,W}\left[\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{A}^\top + \boldsymbol{A}\boldsymbol{x}\boldsymbol{b}^\top + \boldsymbol{A}\boldsymbol{x}\boldsymbol{w}^\top + \boldsymbol{b}\boldsymbol{x}^\top\boldsymbol{A}^\top + \boldsymbol{b}\boldsymbol{b}^\top + \boldsymbol{b}\boldsymbol{w}^\top\right. \\
&\quad \left. + \boldsymbol{w}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})^\top + \boldsymbol{w}\boldsymbol{w}^\top)\right] - \boldsymbol{\mu}_y\boldsymbol{\mu}_y^\top .
\end{aligned}$$

We use the linearity of the expected value, move all constants out of the expected value, and exploit the independence of $\boldsymbol{w}$ and $\boldsymbol{x}$:

$$\begin{aligned}
\boldsymbol{\Sigma}_y &= \boldsymbol{A}\mathbb{E}_X[\boldsymbol{x}\boldsymbol{x}^\top]\boldsymbol{A}^\top + \boldsymbol{A}\mathbb{E}_X[\boldsymbol{x}]\boldsymbol{b}^\top + \boldsymbol{b}\mathbb{E}_X[\boldsymbol{x}^\top]\boldsymbol{A}^\top + \boldsymbol{b}\boldsymbol{b}^\top + \mathbb{E}_W[\boldsymbol{w}\boldsymbol{w}^\top] \\
&\quad - (\boldsymbol{A}\boldsymbol{\mu}_x + \boldsymbol{b})(\boldsymbol{A}\boldsymbol{\mu}_x + \boldsymbol{b})^\top ,
\end{aligned}$$

where we used our previous result for $\boldsymbol{\mu}_y$. Note that $\mathbb{E}_W[\boldsymbol{w}] = \boldsymbol{0}$. We continue as follows:

$$\begin{aligned}
\boldsymbol{\Sigma}_y &= \boldsymbol{A}\mathbb{E}_X[\boldsymbol{x}\boldsymbol{x}^\top]\boldsymbol{A}^\top + \boldsymbol{A}\boldsymbol{\mu}_x\boldsymbol{b}^\top + \boldsymbol{b}\boldsymbol{\mu}_x^\top\boldsymbol{A}^\top + \boldsymbol{b}\boldsymbol{b}^\top + \boldsymbol{Q} \\
&\quad - \boldsymbol{A}\boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top\boldsymbol{A}^\top - \boldsymbol{A}\boldsymbol{\mu}_x\boldsymbol{b}^\top - \boldsymbol{b}\boldsymbol{\mu}_x^\top\boldsymbol{A}^\top - \boldsymbol{b}\boldsymbol{b}^\top \\
&= \boldsymbol{A}\underbrace{\left(\mathbb{E}_X[\boldsymbol{x}\boldsymbol{x}^\top] - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top\right)}_{=\boldsymbol{\Sigma}_x}\boldsymbol{A}^\top + \boldsymbol{Q} \\
&= \boldsymbol{A}\boldsymbol{\Sigma}_x\boldsymbol{A}^\top + \boldsymbol{Q} .
\end{aligned}$$

Alternatively, we could have exploited

$$\begin{aligned}
\mathbb{V}_{\boldsymbol{y}}[\boldsymbol{y}] &= \mathbb{V}_{\boldsymbol{x},\boldsymbol{w}}[\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{w}] \overset{\text{i.i.d.}}{=} \mathbb{V}_{\boldsymbol{x}}[\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}] + \mathbb{V}_{\boldsymbol{w}}[\boldsymbol{w}] \\
&= \boldsymbol{A}\mathbb{V}_{\boldsymbol{x}}\boldsymbol{A}^\top + \boldsymbol{Q} = \boldsymbol{A}\boldsymbol{\Sigma}_x\boldsymbol{A}^\top + \boldsymbol{Q} .
\end{aligned}$$

c. The random variable $\boldsymbol{y}$ is being transformed according to the measurement mapping

$$\boldsymbol{z} = \boldsymbol{C}\boldsymbol{y} + \boldsymbol{v} ,$$

where $\boldsymbol{z} \in \mathbb{R}^F$, $\boldsymbol{C} \in \mathbb{R}^{F\times E}$, and $\boldsymbol{v} \sim \mathcal{N}\big(\boldsymbol{v}\,|\,\boldsymbol{0},\, \boldsymbol{R}\big)$ is independent Gaussian (measurement) noise.

- Write down $p(\boldsymbol{z}\,|\,\boldsymbol{y})$.

$$\begin{aligned}
p(\boldsymbol{z}\,|\,\boldsymbol{y}) &= \mathcal{N}\big(\boldsymbol{z}\,|\,\boldsymbol{C}\boldsymbol{y},\, \boldsymbol{R}\big) \\
&= (2\pi)^{-F/2}\,|\boldsymbol{R}|^{-1/2}\exp\big(-\tfrac{1}{2}(\boldsymbol{z} - \boldsymbol{C}\boldsymbol{y})^\top\boldsymbol{R}^{-1}(\boldsymbol{z} - \boldsymbol{C}\boldsymbol{y})\big) .
\end{aligned}$$

- Compute $p(\boldsymbol{z})$, i.e., the mean $\boldsymbol{\mu}_z$ and the covariance $\boldsymbol{\Sigma}_z$. Derive your result in detail.

  Remark: Since $\boldsymbol{y}$ is Gaussian and $\boldsymbol{z}$ is a linear transformation of $\boldsymbol{y}$, $p(\boldsymbol{z})$ is Gaussian, too. Let's compute its moments (similar to the "time update"):

$$\begin{aligned}\boldsymbol{\mu}_z &= \mathbb{E}_Z[\boldsymbol{z}] = \mathbb{E}_{Y,V}[\boldsymbol{C}\boldsymbol{y} + \boldsymbol{v}] = \boldsymbol{C}\mathbb{E}_Y[\boldsymbol{y}] + \underbrace{\mathbb{E}_V[\boldsymbol{v}]}_{=\boldsymbol{0}} \\ &= \boldsymbol{C}\boldsymbol{\mu}_y\,.\end{aligned}$$

  For the covariance matrix, we compute

$$\begin{aligned}\boldsymbol{\Sigma}_z &= \mathbb{E}_Z[\boldsymbol{z}\boldsymbol{z}^\top] - \mathbb{E}_Z[\boldsymbol{z}]\mathbb{E}_Z[\boldsymbol{z}]^\top \\ &= \mathbb{E}_{Y,V}\left[(\boldsymbol{C}\boldsymbol{y} + \boldsymbol{v})(\boldsymbol{C}\boldsymbol{y} + \boldsymbol{v})^\top\right] - \boldsymbol{\mu}_z\boldsymbol{\mu}_z^\top \\ &= \mathbb{E}_{Y,V}\left[\boldsymbol{C}\boldsymbol{y}\boldsymbol{y}^\top\boldsymbol{C}^\top + \boldsymbol{C}\boldsymbol{y}\boldsymbol{v}^\top + \boldsymbol{v}(\boldsymbol{C}\boldsymbol{y})^\top + \boldsymbol{v}\boldsymbol{v}^\top)\right] - \boldsymbol{\mu}_z\boldsymbol{\mu}_z^\top\,.\end{aligned}$$

  We use the linearity of the expected value, move all constants out of the expected value, and exploit the independence of $\boldsymbol{v}$ and $\boldsymbol{y}$:

$$\boldsymbol{\Sigma}_z = \boldsymbol{C}\mathbb{E}_Y[\boldsymbol{y}\boldsymbol{y}^\top]\boldsymbol{C}^\top + \mathbb{E}_V[\boldsymbol{v}\boldsymbol{v}^\top] - \boldsymbol{C}\boldsymbol{\mu}_y\boldsymbol{\mu}_y^\top\boldsymbol{C}^\top\,,$$

  where we used our previous result for $\boldsymbol{\mu}_z$. Note that $\mathbb{E}_V[\boldsymbol{v}] = \boldsymbol{0}$. We continue as follows:

$$\begin{aligned}\boldsymbol{\Sigma}_y &= \boldsymbol{C}\underbrace{\left(\mathbb{E}_Y[\boldsymbol{y}\boldsymbol{y}^\top] - \boldsymbol{\mu}_y\boldsymbol{\mu}_y^\top\right)}_{=\boldsymbol{\Sigma}_x}\boldsymbol{C}^\top + \boldsymbol{R} \\ &= \boldsymbol{C}\boldsymbol{\Sigma}_y\boldsymbol{C}^\top + \boldsymbol{R}\,.\end{aligned}$$

d. Now, a value $\hat{\boldsymbol{y}}$ is measured. Compute the posterior distribution $p(\boldsymbol{x}\,|\,\hat{\boldsymbol{y}})$.
Hint for solution: This posterior is also Gaussian, i.e., we need to determine only its mean and covariance matrix. Start by explicitly computing the joint Gaussian $p(\boldsymbol{x}, \boldsymbol{y})$. This also requires us to compute the cross-covariances $\mathrm{Cov}_{\boldsymbol{x},\boldsymbol{y}}[\boldsymbol{x}, \boldsymbol{y}]$ and $\mathrm{Cov}_{\boldsymbol{y},\boldsymbol{x}}[\boldsymbol{y}, \boldsymbol{x}]$. Then apply the rules for Gaussian conditioning.
We derive the posterior distribution following the second hint since we do not have to worry about normalization constants:
Assume, we know the joint Gaussian distribution

$$p(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} \,\middle|\, \begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xy}\\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_y\end{bmatrix}\right), \tag{6.11}$$

where we defined $\boldsymbol{\Sigma}_{xy} := \mathrm{Cov}_{\boldsymbol{x},\boldsymbol{y}}[\boldsymbol{x}, \boldsymbol{y}]$.
Now, we apply the rules for Gaussian conditioning to obtain

$$\begin{aligned}p(\boldsymbol{x}\,|\,\hat{\boldsymbol{y}}) &= \mathcal{N}\big(\boldsymbol{x}\,|\,\boldsymbol{\mu}_{x\,|\,y}, \boldsymbol{\Sigma}_{x\,|\,y}\big) \\ \boldsymbol{\mu}_{x\,|\,y} &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_y^{-1}(\hat{\boldsymbol{y}} - \boldsymbol{\mu}_y) \\ \boldsymbol{\Sigma}_{x\,|\,y} &= \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{yx}\,.\end{aligned}$$

Looking at (6.11), it remains to compute the cross-covariance term $\boldsymbol{\Sigma}_{xy}$ (the marginal distributions $p(\boldsymbol{x})$ and $p(\boldsymbol{y})$ are known and $\boldsymbol{\Sigma}_{yx} = \boldsymbol{\Sigma}_{xy}^\top$):

$$\boldsymbol{\Sigma}_{xy} = \mathrm{Cov}[\boldsymbol{x}, \boldsymbol{y}] = \mathbb{E}_{X,Y}[\boldsymbol{x}\boldsymbol{y}^\top] - \mathbb{E}_X[\boldsymbol{x}]\mathbb{E}_Y[\boldsymbol{y}]^\top = \mathbb{E}_{X,Y}[\boldsymbol{x}\boldsymbol{y}^\top] - \boldsymbol{\mu}_x\boldsymbol{\mu}_y^\top\,,$$

where $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ are known. Hence, it remains to compute

$$\begin{aligned}\mathbb{E}_{X,Y}[\boldsymbol{x}\boldsymbol{y}^\top] &= \mathbb{E}_X[\boldsymbol{x}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}+\boldsymbol{w})^\top] = \mathbb{E}_X[\boldsymbol{x}\boldsymbol{x}^\top]\boldsymbol{A}^\top + \mathbb{E}_X[\boldsymbol{x}]\boldsymbol{b}^\top \\ &= \mathbb{E}_X[\boldsymbol{x}\boldsymbol{x}^\top]\boldsymbol{A}^\top + \boldsymbol{\mu}_x\boldsymbol{b}^\top\,.\end{aligned}$$

With

$$\boldsymbol{\mu}_x\boldsymbol{\mu}_y^\top = \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top\boldsymbol{A}^\top + \boldsymbol{\mu}_x\boldsymbol{b}^\top$$

we obtain the desired cross-covariance

$$\begin{aligned}\boldsymbol{\Sigma}_{xy} &= \mathbb{E}_X[\boldsymbol{x}\boldsymbol{x}^\top]\boldsymbol{A}^\top + \boldsymbol{\mu}_x\boldsymbol{b}^\top - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top\boldsymbol{A}^\top - \boldsymbol{\mu}_x\boldsymbol{b}^\top \\ &= \big(\mathbb{E}_X[\boldsymbol{x}\boldsymbol{x}^\top] - \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top\big)\boldsymbol{A}^\top = \boldsymbol{\Sigma}_x\boldsymbol{A}^\top\,.\end{aligned}$$

And finally

$$\begin{aligned}\boldsymbol{\mu}_{x\,|\,y} &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_x\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{\Sigma}_x\boldsymbol{A}^\top + \boldsymbol{Q})^{-1}(\hat{\boldsymbol{y}} - \boldsymbol{A}\boldsymbol{\mu}_x - \boldsymbol{b})\,, \\ \boldsymbol{\Sigma}_{x\,|\,y} &= \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x\boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{\Sigma}_x\boldsymbol{A}^\top + \boldsymbol{Q})^{-1}\boldsymbol{A}\boldsymbol{\Sigma}_x\,.\end{aligned}$$

6.13 **Probability Integral Transformation**
Given a continuous random variable x, with cdf $F_x(x)$, show that the random variable $y = F_x(x)$ is uniformly distributed.

Proof We need to show that the cumulative distribution function of Y defines a distribution of a uniform random variable. Recall that by the axioms of probability (Section 6.1) probabilities must be non-negative and sum/integrate to one. Therefore, the range of possible values of $Y = F_X(x)$ is the interval $[0, 1]$. For any $F_X(x)$, the inverse $F_X^{-1}(x)$ exists because we assumed that $F_X(x)$ is strictly monotonically increasing, which we will use in the following.
Given any continuous random variable X, the definition of a cdf gives

$$\begin{aligned}F_Y(y) &= P(Y \leqslant y) \\ &= P(F_X(x) \leqslant y) && \text{transformation of interest} \\ &= P(X \leqslant F_X^{-1}(y)) && \text{inverse exists} \\ &= F_X(F_X^{-1}(y)) && \text{definition of cdf} \\ &= y\,,\end{aligned}$$

where the last line is due to the fact that $F_X(x)$ composed with its inverse results in an identity transformation. The statement $F_Y(y) = y$ along with the fact that y lies in the interval $[0, 1]$ means that $F_Y(x)$ is the cdf of the uniform random variable on the unit interval. □

7

Continuous Optimization

Exercises

7.1 Consider the univariate function

$$f(x) = x^3 + 6x^2 - 3x - 5.$$

Find its stationary points and indicate whether they are maximum, minimum, or saddle points.

Given the function $f(x)$, we obtain the following gradient and Hessian,

$$\frac{\mathrm{d}f}{\mathrm{d}x} = 3x^2 + 12x - 3$$
$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} = 6x + 12\,.$$

We find stationary points by setting the gradient to zero, and solving for x. One option is to use the formula for quadratic functions, but below we show how to solve it using completing the square. Observe that

$$(x+2)^2 = x^2 + 4x + 4$$

and therefore (after dividing all terms by 3),

$$(x^2 + 4x) - 1 = ((x+2)^2 - 4) - 1\,.$$

By setting this to zero, we obtain that

$$(x+2)^2 = 5,$$

and hence

$$x = -2 \pm \sqrt{5}.$$

Substituting the solutions of $\frac{\mathrm{d}f}{\mathrm{d}x} = 0$ into the Hessian gives

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(-2-\sqrt{5}) \approx -13.4 \quad \text{and} \quad \frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(-2+\sqrt{5}) \approx 13.4\,.$$

This means that the left stationary point is a maximum and the right one is a minimum. See Figure 7.1 for an illustration.

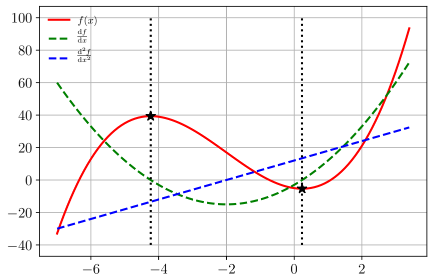
7.2 Consider the update equation for stochastic gradient descent (Equation (7.15)). Write down the update when we use a mini-batch size of one.

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \gamma_i (\nabla L(\boldsymbol{\theta}_i))^\top = \boldsymbol{\theta}_i - \gamma_i \nabla L_k(\boldsymbol{\theta}_i)^\top$$

where k is the index of the example that is randomly chosen.

7.3 Consider whether the following statements are true or false:

1 A plot of
on $f(x)$
n its
nd



a. The intersection of any two convex sets is convex.
b. The union of any two convex sets is convex.
c. The difference of a convex set A from another convex set B is convex.

a. true
b. false
c. false

7.4 Consider whether the following statements are true or false:

a. The sum of any two convex functions is convex.
b. The difference of any two convex functions is convex.
c. The product of any two convex functions is convex.
d. The maximum of any two convex functions is convex.

a. true
b. false
c. false
d. true

7.5 Express the following optimization problem as a standard linear program in matrix notation

$$\max_{\boldsymbol{x}\in\mathbb{R}^2,\ \xi\in\mathbb{R}} \boldsymbol{p}^\top \boldsymbol{x} + \xi$$

subject to the constraints that $\xi \geqslant 0$, $x_0 \leqslant 0$ and $x_1 \leqslant 3$.
The optimization target and constraints must all be linear. To make the optimization target linear, we combine $\boldsymbol{x}$ and ξ into one matrix equation:

$$\max_{\boldsymbol{x}\in\mathbb{R}^2,\xi\in\mathbb{R}} \begin{bmatrix} p_0 \\ p_1 \\ 1 \end{bmatrix}^\top \begin{bmatrix} x_0 \\ x_1 \\ \xi \end{bmatrix}$$

We can then combine all of the inequalities into one matrix inequality:

$$\text{subject to } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \xi \end{bmatrix} - \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} \leqslant \boldsymbol{0}$$

7.6 Consider the linear program illustrated in Figure 7.9,

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} - \begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\text{subject to} \quad \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leqslant \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix}$$

Derive the dual linear program using Lagrange duality.

Write down the Lagrangian:

$$\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = -\begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top \left(\begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} \boldsymbol{x} - \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix} \right)$$

Rearrange and factorize $\boldsymbol{x}$:

$$\mathfrak{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \left(-\begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top + \boldsymbol{\lambda}^\top \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} \right) \boldsymbol{x} - \boldsymbol{\lambda}^\top \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix}$$

Differentiate with respect to $\boldsymbol{x}$ and set to zero:

$$\nabla_{\boldsymbol{x}} \mathfrak{L} = -\begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top + \boldsymbol{\lambda}^\top \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} = 0$$

Then substitute back into the Lagrangian to obtain the dual Lagrangian:

$$\mathfrak{D}(\boldsymbol{\lambda}) = -\boldsymbol{\lambda}^\top \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix}$$

The dual optimization problem is therefore

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \quad -\boldsymbol{\lambda}^\top \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix}$$

$$\text{subject to} \quad -\begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top + \boldsymbol{\lambda}^\top \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} = 0$$

$$\boldsymbol{\lambda} \geqslant \boldsymbol{0}.$$

7.7 Consider the quadratic program illustrated in Figure 7.4,

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \frac{1}{2}\begin{bmatrix}x_1\\x_2\end{bmatrix}^\top \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} + \begin{bmatrix}5\\3\end{bmatrix}^\top \begin{bmatrix}x_1\\x_2\end{bmatrix}$$

$$\text{subject to} \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} \leqslant \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

Derive the dual quadratic program using Lagrange duality.

Let $\boldsymbol{Q} = \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}$, $\boldsymbol{c} = \begin{bmatrix}5\\3\end{bmatrix}$, $\boldsymbol{A} = \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}$, and $\boldsymbol{b} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$. Then by (7.45) and (7.52) the dual optimization problem is

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^4} -\frac{1}{2}\left(\begin{bmatrix}5\\3\end{bmatrix} + \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}^\top \boldsymbol{\lambda}\right)^\top \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}^{-1}\left(\begin{bmatrix}5\\3\end{bmatrix} + \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}^\top \boldsymbol{\lambda}\right) - \boldsymbol{\lambda}^\top \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

subject to $\boldsymbol{\lambda} \geqslant \boldsymbol{0}$.

which expands to

$$\max_{\boldsymbol{\lambda}\in\mathbb{R}^4} -\frac{1}{14}\left(88 + \begin{bmatrix}33\\-35\\1\\-3\end{bmatrix}^\top \boldsymbol{\lambda} + \boldsymbol{\lambda}^\top \begin{bmatrix}4 & -4 & -1 & 1\\-4 & 4 & 1 & -1\\-1 & 1 & 2 & -2\\1 & -1 & -2 & 2\end{bmatrix}\boldsymbol{\lambda}\right)$$

subject to $\boldsymbol{\lambda} \geqslant \boldsymbol{0}$.

Alternative derivation

Write down the Lagrangian:

$$\mathfrak{L}(\boldsymbol{x},\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^\top \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}\boldsymbol{x} + \begin{bmatrix}5\\3\end{bmatrix}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top\left(\begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}\boldsymbol{x} - \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\right)$$

Differentiate with respect to $\boldsymbol{x}$ and set to zero:

$$\nabla_{\boldsymbol{x}}\mathfrak{L} = \boldsymbol{x}^\top \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix} + \begin{bmatrix}5\\3\end{bmatrix}^\top + \boldsymbol{\lambda}^\top \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix} = 0$$

Solve for $\boldsymbol{x}$:

$$\boldsymbol{x}^\top = -\left(\begin{bmatrix}5\\3\end{bmatrix}^\top + \boldsymbol{\lambda}^\top \begin{bmatrix}1 & 0\\-1 & 0\\0 & 1\\0 & -1\end{bmatrix}\right)\begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}^{-1}$$

Then substitute back to get the dual Lagrangian:

$$\mathfrak{D}(\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^\top \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 5 \\ 3 \end{bmatrix}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top \left(\begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \boldsymbol{x} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right)$$

7.8 Consider the following convex optimization problem

$$\begin{aligned} \min_{\boldsymbol{w} \in \mathbb{R}^D} \quad & \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} \\ \text{subject to} \quad & \boldsymbol{w}^\top \boldsymbol{x} \geqslant 1\,. \end{aligned}$$

Derive the Lagrangian dual by introducing the Lagrange multiplier λ.
First we express the convex optimization problem in standard form,

$$\begin{aligned} &\min_{\boldsymbol{w} \in \mathbb{R}^D} \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} \\ &\text{subject to } 1 - \boldsymbol{w}^\top \boldsymbol{x} \leqslant 0\,. \end{aligned}$$

By introducing a Lagrange multiplier $\lambda \geqslant 0$, we obtain the following Lagrangian

$$\mathfrak{L}(\boldsymbol{w}) = \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} + \lambda(1 - \boldsymbol{w}^\top \boldsymbol{x})$$

Taking the gradient of the Lagrangian with respect to $\boldsymbol{w}$ gives

$$\frac{\mathrm{d}\mathfrak{L}(\boldsymbol{w})}{\mathrm{d}\boldsymbol{w}} = \boldsymbol{w}^\top - \lambda \boldsymbol{x}^\top\,.$$

Setting the gradient to zero and solving for $\boldsymbol{w}$ gives

$$\boldsymbol{w} = \lambda \boldsymbol{x}\,.$$

Substituting back into $\mathfrak{L}(\boldsymbol{w})$ gives the dual Lagrangian

$$\begin{aligned} \mathfrak{D}(\lambda) &= \frac{\lambda^2}{2}\boldsymbol{x}^\top \boldsymbol{x} + \lambda - \lambda^2 \boldsymbol{x}^\top \boldsymbol{x} \\ &= -\frac{\lambda^2}{2}\boldsymbol{x}^\top \boldsymbol{x} + \lambda\,. \end{aligned}$$

Therefore the dual optimization problem is given by

$$\begin{aligned} &\max_{\lambda \in \mathbb{R}} -\frac{\lambda^2}{2}\boldsymbol{x}^\top \boldsymbol{x} + \lambda \\ &\text{subject to } \lambda \geqslant 0\,. \end{aligned}$$

7.9 Consider the negative entropy of $\boldsymbol{x} \in \mathbb{R}^D$,

$$f(\boldsymbol{x}) = \sum_{d=1}^{D} x_d \log x_d\,.$$

Derive the convex conjugate function $f^*(s)$, by assuming the standard dot

product.
Hint: Take the gradient of an appropriate function and set the gradient to zero.
From the definition of the Legendre Fenchel conjugate

$$f^*(\boldsymbol{s}) = \sup_{\boldsymbol{x} \in \mathbb{R}^D} \sum_{d=1}^{D} s_d x_d - x_d \log x_d\,.$$

Define a function (for notational convenience)

$$g(\boldsymbol{x}) = \sum_{d=1}^{D} s_d x_d - x_d \log x_d\,.$$

The gradient of $g(\boldsymbol{x})$ with respect to x_d is

$$\begin{aligned} \frac{\mathrm{d}g(\boldsymbol{x})}{\mathrm{d}x_d} &= s_d - \frac{x_d}{x_d} - \log x_d \\ &= s_d - 1 - \log x_d\,. \end{aligned}$$

Setting the gradient to zero gives

$$x_d = \exp(s_d - 1)\,.$$

Substituting the optimum value of x_d back into $f^*(\boldsymbol{s})$ gives

$$\begin{aligned} f^*(\boldsymbol{s}) &= \sum_{d=1}^{D} = s_d \exp(s_d - 1) - (s_d - 1)\exp(s_d - 1) \\ &= \exp(s_d - 1)\,. \end{aligned}$$

7.10 Consider the function

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c\,,$$

where $\boldsymbol{A}$ is strictly positive definite, which means that it is invertible. Derive the convex conjugate of $f(\boldsymbol{x})$.
Hint: Take the gradient of an appropriate function and set the gradient to zero.
From the definition of the Legendre Fenchel transform,

$$f^*(\boldsymbol{s}) = \sup_{\boldsymbol{x} \in \mathbb{R}^D} \boldsymbol{s}^\top \boldsymbol{x} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x} - c\,.$$

Define a function (for notational convenience)

$$g(\boldsymbol{x}) = \boldsymbol{s}^\top \boldsymbol{x} - \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x} - c\,.$$

The gradient of $g(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ is

$$\frac{\mathrm{d}g(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} = \boldsymbol{s}^\top - \boldsymbol{x}^\top \boldsymbol{A} - \boldsymbol{b}^\top\,.$$

Setting the gradient to zero gives (note that $\boldsymbol{A}^\top = \boldsymbol{A}$)

$$\begin{aligned} \boldsymbol{s}\boldsymbol{A}^\top \boldsymbol{x} - \boldsymbol{b} &= 0 \\ \boldsymbol{A}^\top \boldsymbol{x} &= \boldsymbol{s} - \boldsymbol{b} \\ \boldsymbol{x} &= \boldsymbol{A}^{-1}(\boldsymbol{s} - \boldsymbol{b})\,. \end{aligned}$$

Substituting the optimum value of $\boldsymbol{x}$ back into $f^*(\boldsymbol{s})$ gives

$$\begin{aligned} f^*(\boldsymbol{s}) &= \boldsymbol{s}^\top \boldsymbol{A}^{-1}(\boldsymbol{s}-\boldsymbol{b}) - \frac{1}{2}(\boldsymbol{s}-\boldsymbol{b})\boldsymbol{A}^{-1}(\boldsymbol{s}-\boldsymbol{b}) - \boldsymbol{b}^\top \boldsymbol{A}^{-1}(\boldsymbol{s}-\boldsymbol{b}) - c \\ &= \frac{1}{2}(\boldsymbol{s}-\boldsymbol{b})^\top \boldsymbol{A}^{-1}(\boldsymbol{s}-\boldsymbol{b}) - c\,. \end{aligned}$$

7.11 The hinge loss (which is the loss used by the support vector machine) is given by

$$L(\alpha) = \max\{0, 1-\alpha\}\,,$$

If we are interested in applying gradient methods such as L-BFGS, and do not want to resort to subgradient methods, we need to smooth the kink in the hinge loss. Compute the convex conjugate of the hinge loss $L^*(\beta)$ where β is the dual variable. Add a ℓ_2 proximal term, and compute the conjugate of the resulting function

$$L^*(\beta) + \frac{\gamma}{2}\beta^2\,,$$

where γ is a given hyperparameter.
Recall that the hinge loss is given by

$$L(\alpha) = \max\{0, 1-\alpha\}$$

The convex conjugate of $L(\alpha)$ is

$$\begin{aligned} L^*(\beta) &= \sup_{\alpha\in\mathbb{R}} \{\alpha\beta - \max\{0, 1-\alpha\}\} \\ &= \begin{cases} \beta & \text{if} \quad -1 \leqslant \beta \leqslant 0, \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

The smoothed conjugate is

$$L^*_\gamma(\beta) = L^*(\beta) + \frac{\gamma}{2}\beta^2.$$

The corresponding primal smooth hinge loss is given by

$$\begin{aligned} L_\gamma(\alpha) &= \sup_{-1\leqslant\beta\leqslant 0} \left\{\alpha\beta - \beta - \frac{\gamma}{2}\beta^2\right\} \\ &= \begin{cases} 1-\alpha-\frac{\gamma}{2} & \text{if} \quad \alpha < 1-\gamma, \\ \frac{(\alpha-1)^2}{2\gamma} & \text{if} \quad 1-\gamma \leqslant \alpha \leqslant 1, \\ 0 & \text{if} \quad \alpha > 1. \end{cases} \end{aligned}$$

$L_\gamma(\alpha)$ is convex and differentiable with the derivative

$$L'_\gamma(\alpha) = \begin{cases} -1 & \text{if} \quad \alpha < 1-\gamma, \\ \frac{\alpha-1}{\gamma} & \text{if} \quad 1-\gamma \leqslant \alpha \leqslant 1, \\ 0 & \text{if} \quad \alpha > 1. \end{cases}$$