

Chapter 2: Linear Algebra - Exercises

2.1

a. We are going to show that $(\mathbb{R} \setminus \{-1\}, \star)$ with the operator defined as

$$a \star b := ab + a + b, \quad a, b \in \mathbb{R} \setminus \{-1\}$$

is an Abelian group. We need to verify that it satisfies all the necessary properties for being a standard group, then verify that the operation is also commutative.

1. **Closure.** For any $x, y \in \mathbb{R} \setminus \{-1\}$, then $x \star y \in \mathbb{R} \setminus \{-1\}$. Clearly,

$$x \star y = xy + x + y \in \mathbb{R},$$

but we must also make sure that no two values give -1 since this is not included in the set. First, we rewrite the expression:

$$x \star y = xy + x + y = (x + 1)(y + 1) - 1$$

Setting this equal to -1 .

$$(x + 1)(y + 1) - 1 = -1$$

Rewriting:

$$(x + 1)(y + 1) = 0$$

Recalling that neither x nor y can be -1 , we can see from this equation that we will never get 0. This shows that $\mathbb{R} \setminus \{-1\}$ is closed under \star .

2. **Associativity.** Assuming $x, y, z \in \mathbb{R} \setminus \{-1\}$ and we are going to show that

$$(x \star y) \star z = x \star (y \star z).$$

Left hand side:

$$\begin{aligned} (x \star y) \star z &= (xy + x + y) \star z \\ &= (xy + x + y)z + (xy + x + y) + z \\ &= zxy + zx + zy + xy + x + y + z \\ &= xyz + xy + xz + yz + x + y + z \end{aligned}$$

Right hand side:

$$\begin{aligned} x \star (y \star z) &= x \star (yz + y + z) \\ &= x(yz + y + z) + x + (yz + y + z) \\ &= xyz + xy + xz + x + yz + y + z \\ &= xyz + xy + xz + yz + x + y + z \end{aligned}$$

Both expressions end up equal, proving associativity.

3. **Neutral element.** For any $x \in \mathbb{R} \setminus \{-1\}$, we want to find some $e \in \mathbb{R} \setminus \{-1\}$ such that

$$x \star e = xe + x + e = x$$

$$e \star x = ex + e + x = x$$

Both are satisfied when $e = 0 \in \mathbb{R} \setminus \{-1\}$ which is the neutral element.

4. **Inverse element.** For any $x \in \mathbb{R} \setminus \{-1\}$ we want to find some inverse $i \in \mathbb{R} \setminus \{-1\}$ such that $x \star i = 0$.

$$x \star i = xi + x + i = 0$$

Collecting i terms and factoring out.

$$(x + 1)i + x = 0$$

Solving for i .

$$i = -\frac{x}{x + 1}$$

This is how the inverse is calculated. It is not defined for $x = -1$ but that is not in the set.

5. **Commutativity.** We also need this property to verify that we have an Abelian group. Let $x, y \in \mathbb{R} \setminus \{-1\}$. Then:

$$x \star y = xy + x + y = yx + y + x = y \star x.$$

We have verified all properties, so we can conclude that $(\mathbb{R} \setminus \{-1\}, \star)$ is an Abelian group.

b. Solving the equation $3 \star x \star x = 15$. First we calculate the left side.

$$\begin{aligned} 3 \star x \star x &= 3 \star (xx + x + x) \\ &= 3 \star (x^2 + 2x) \\ &= 3(x^2 + 2x) + 3 + x^2 + 2x \\ &= 3x^2 + 6x + 3 + x^2 + 2x \\ &= 4x^2 + 8x + 3 \end{aligned}$$

We have

$$4x^2 + 8x + 3 = 15,$$

which is also

$$4x^2 + 8x - 12 = 0,$$

which can be further simplified into

$$x^2 + 2x - 3 = 0.$$

By using the quadratic formula, it can be shown that $x = 1$ or $x = -3$. These solutions can be verified by replacing the values for x in the original equation.

Checking $x = -3$.

$$\begin{aligned}3 \star x \star x &= 3 \star (-3) \star (-3) \\&= 3 \star ((-3)(-3) + (-3) + (-3)) \\&= 3 \star (9 - 6) \\&= 3 \star 3 \\&= 3 \cdot 3 + 3 + 3 \\&= 9 + 6 \\&= 15\end{aligned}$$

Checking $x = 1$.

$$\begin{aligned}3 \star x \star x &= 3 \star 1 \star 1 \\&= 3 \star (1 \cdot 1 + 1 + 1) \\&= 3 \star 3 \\&= 3 \cdot 3 + 3 + 3 \\&= 9 + 6 \\&= 15\end{aligned}$$

2.2 - Skipped

2.3

We have the following set defined

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

and the operator \cdot is defined as the standard matrix multiplication. We will first determine if (\mathcal{G}, \cdot) is a group. We check the properties.

1. Closure.

We will take two arbitrary matrices from \mathcal{G} , multiply them, and check if their product qualifies as a member of \mathcal{G} . Since we are multiplying two 3×3 matrices, the result will be a 3×3 matrix.

Assuming $a, b, c \in \mathbb{R}$ and $x, y, z \in \mathbb{R}$.

$$\begin{aligned} \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} &= \begin{bmatrix} 1(1) + a(0) + c(0) & 1(x) + a(1) + c(0) & 1(z) + a(y) + c(1) \\ 0(1) + 1(0) + b(0) & 0(x) + 1(1) + b(0) & 0(z) + 1(y) + b(1) \\ 0(1) + 0(0) + 1(0) & 0(x) + 0(1) + 1(0) & 0(z) + 0(y) + 1(1) \end{bmatrix} \\ &= \begin{bmatrix} 1 + 0 + 0 & x + a + 0 & z + ay + c \\ 0 + 0 + 0 & 0 + 1 + 0 & 0 + y + b \\ 0 + 0 + 0 & 0 + 0 + 0 & 0 + 0 + 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & x + a & z + ay + c \\ 0 & 1 & y + b \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

Since $x + a, y + b, z + ay + c \in \mathbb{R}$, this matrix is in \mathcal{G} which shows we have closure.

2. Associativity.

For matrices $A, B, C \in \mathcal{G}$ we have to show that $(AB)C = A(BC)$. This can be done explicitly by multiplying three matrices in both ways and comparing the results. Assuming $a_1, a_2, a_3 \in \mathbb{R}$, $b_1, b_2, b_3 \in \mathbb{R}$ and $c_1, c_2, c_3 \in \mathbb{R}$, we define the following three matrices.

$$A = \begin{bmatrix} 1 & a_1 & a_3 \\ 0 & 1 & a_2 \\ 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & b_1 & b_3 \\ 0 & 1 & b_2 \\ 0 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & c_1 & c_3 \\ 0 & 1 & c_2 \\ 0 & 0 & 1 \end{bmatrix}$$

Checking the first multiplication.

$$(AB)C = \begin{bmatrix} 1 & a_1 + b_1 & a_3 + a_1b_2 + b_3 \\ 0 & 1 & a_2 + b_2 \\ 0 & 0 & 1 \end{bmatrix} C = \begin{bmatrix} 1 & a_1 + b_1 + c_1 & a_3 + b_3 + c_3 + a_1b_2 + a_1c_2 + b_1c_2 \\ 0 & 1 & a_2 + b_2 + c_2 \\ 0 & 0 & 1 \end{bmatrix}$$

Checking the second multiplication.

$$A(BC) = A \begin{bmatrix} 1 & b_1 + c_1 & b_3 + b_1c_2 + c_3 \\ 0 & 1 & b_2 + c_2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & a_1 + b_1 + c_1 & a_3 + b_3 + c_3 + a_1b_2 + a_1c_2 + b_1c_2 \\ 0 & 1 & a_2 + b_2 + c_2 \\ 0 & 0 & 1 \end{bmatrix}$$

The matrices are the same and are members of \mathcal{G} (which follows from the closure property). This shows that associativity is satisfied.

3. Neutral element.

We are working with 3×3 matrices, and the standard I_3 identity matrix is the neutral element. For any $A \in \mathcal{G}$, then $I_3 A = A I_3 = A$. Since $0 \in \mathbb{R}$, we have $I_3 \in \mathcal{G}$. (That is, we can set $x = y = z = 0$ in the definition and see that the identity matrix satisfies the conditions for membership in \mathcal{G} .)

4. Inverse element.

We need to find an inverse to a matrix $A \in \mathcal{G}$ and check if it is in \mathcal{G} . We assume that $a_1, a_2, a_3 \in \mathbb{R}$. Without writing out all the calculations, we can show that:

$$\begin{bmatrix} 1 & a_1 & a_3 \\ 0 & 1 & a_2 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -a_1 & a_1 a_2 - a_3 \\ 0 & 1 & -a_2 \\ 0 & 0 & 1 \end{bmatrix}$$

Since $-a_1, -a_2, a_1 a_2 - a_3 \in \mathbb{R}$ the inverse matrix $A^{-1} \in \mathcal{G}$. Hence the inverse element condition is satisfied and we can conclude that (\mathcal{G}, \cdot) is a group.

Finally we check if the commutativity property is satisfied to see if the group is also Abelian.

5. Commutativity.

For two matrices $A, B \in \mathcal{G}$ we want to check if $AB = BA$. Assuming $a_1, a_2, a_3 \in \mathbb{R}$ and $b_1, b_2, b_3 \in \mathbb{R}$, and we define the matrices as

$$A = \begin{bmatrix} 1 & a_1 & a_3 \\ 0 & 1 & a_2 \\ 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & b_1 & b_3 \\ 0 & 1 & b_2 \\ 0 & 0 & 1 \end{bmatrix}$$

We have already calculated AB above.

$$AB = \begin{bmatrix} 1 & a_1 + b_1 & a_3 + a_1 b_2 + b_3 \\ 0 & 1 & a_2 + b_2 \\ 0 & 0 & 1 \end{bmatrix}$$

Calculating BA .

$$BA = \begin{bmatrix} 1 & a_1 + b_1 & a_3 + a_2 b_1 + b_3 \\ 0 & 1 & a_2 + b_2 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that these matrices are NOT the same due to the difference in the top right element.

$$a_3 + a_1 b_2 + b_3 \neq a_3 + a_2 b_1 + b_3$$

This violates the commutative property, and therefore (\mathcal{G}, \cdot) is NOT an Abelian group.

2.4 Calculating som matrices - if possible. Recall that a $m \times n$ matrix has m rows and n columns. The **columns** in the first matrix must correspond to the **rows** in the second matrix. In that case: $(m \times n) \times (n \times p) = m \times p$ matrix.

a. This is a 3×2 matrix multiplied by a 3×3 matrix. Not possible.

b.

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

c.

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

d.

$$\begin{aligned} \begin{bmatrix} 1 & 2 & 1 & 2 \\ 4 & 1 & -1 & -4 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 1 & -1 \\ 2 & 1 \\ 5 & 2 \end{bmatrix} &= \begin{bmatrix} 1(0) + 2(1) + 1(2) + 2(5) & 1(3) + 2(-1) + 1(1) + 2(2) \\ 4(0) + 1(1) + (-1)(2) + (-4)(5) & 4(3) + 1(-1) + (-1)(1) + (-4)(2) \end{bmatrix} \\ &= \begin{bmatrix} 0+2+2+10 & 3+-2+1+4 \\ 0+1-2-20 & 12-1-1-8 \end{bmatrix} \\ &= \begin{bmatrix} 14 & 6 \\ -21 & 2 \end{bmatrix} \end{aligned}$$

e.

$$\begin{aligned}
 \begin{bmatrix} 0 & 3 \\ 1 & -1 \\ 2 & 1 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 2 \\ 4 & 1 & -1 & -4 \end{bmatrix} &= \begin{bmatrix} 0(1) + 3(4) & 0(2) + 3(1) & 0(1) + 3(-1) & 0(2) + 3(-4) \\ 1(1) + (-1)(4) & 1(2) + (-1)(1) & 1(1) + (-1)(-1) & 1(2) + (-1)(-4) \\ 2(1) + 1(4) & 2(2) + 1(1) & 2(1) + 1(-1) & 2(2) + 1(-4) \\ 5(1) + 2(4) & 5(2) + 2(1) & 5(1) + 2(-1) & 5(2) + 2(-4) \end{bmatrix} \\
 &= \begin{bmatrix} 0 + 12 & 0 + 3 & 0 - 3 & 0 - 12 \\ 1 - 4 & 2 - 1 & 1 + 1 & 2 + 4 \\ 2 + 4 & 4 + 1 & 2 - 1 & 4 - 4 \\ 5 + 8 & 10 + 2 & 5 - 2 & 10 - 8 \end{bmatrix} \\
 &= \begin{bmatrix} 12 & 3 & -3 & -12 \\ -3 & 1 & 2 & 6 \\ 6 & 5 & 1 & 0 \\ 13 & 12 & 3 & 2 \end{bmatrix}
 \end{aligned}$$

2.5 Finding the set \mathcal{S} of all solutions to the following inhomogeneous linear systems $\mathbf{Ax} = \mathbf{b}$. This is done by augmenting the matrix and using row reduction to find the echelon form.

a.

$$[\mathbf{A} \mid \mathbf{b}] = \left[\begin{array}{cccc|c} 1 & 1 & -1 & -1 & 1 \\ 2 & 5 & -7 & -5 & -2 \\ 2 & -1 & 1 & 3 & 4 \\ 5 & 2 & -4 & 2 & 6 \end{array} \right] \sim \left[\begin{array}{cccc|c} 1 & 0 & 0 & \frac{3}{2} & 0 \\ 0 & 1 & 0 & -\frac{3}{2} & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

We have a $0 = 1$ contradiction in row four. This means the equation has no solutions, so $\mathcal{S} = \emptyset$.

b.

$$[\mathbf{A} \mid \mathbf{b}] = \left[\begin{array}{ccccc|c} 1 & -1 & 0 & 0 & 1 & 3 \\ 1 & 1 & 0 & -3 & 0 & 6 \\ 2 & -1 & 0 & 1 & -1 & 5 \\ -1 & 2 & 0 & -2 & -1 & -1 \end{array} \right] \sim \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & -1 & 3 \\ 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

There are no contradictions, so there is a solution set to this equation. The variable x_3 is not used and x_5 is a free variable. The solution can be expressed as the following linear equation.

$$\begin{aligned}
 x_1 - x_5 &= 3 \\
 x_2 - 2x_5 &= 0 \\
 x_4 - x_5 &= -1 \\
 x_5 &= x_5
 \end{aligned}$$

We can check that this is a solution. If we set $x_5 = 0$ and use $x_3 = 0$, we get the vector:

$$\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix}$$

Let's multiply this to the matrix \mathbf{A} and check if we get the solution \mathbf{b} .

$$\begin{aligned}
 \mathbf{Ax} &= \begin{bmatrix} 1 & -1 & 0 & 0 & 1 \\ 1 & 1 & 0 & -3 & 0 \\ 2 & -1 & 0 & 1 & -1 \\ -1 & 2 & 0 & -2 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1(3) + (-1)(0) + 0(0) + 0(-1) + 1(0) \\ 1(3) + 1(0) + 0(0) + (-3)(-1) + 0(0) \\ 2(3) + (-1)(0) + 0(0) + 1(-1) + (-1)(0) \\ (-1)(3) + 2(0) + 0(0) + (-2)(-1) + (-1)(0) \end{bmatrix} \\
 &= \begin{bmatrix} 3 + 0 + 0 + 0 + 0 \\ 3 + 0 + 0 + 3 + 0 \\ 6 + 0 + 0 - 1 + 0 \\ -3 + 0 + 0 + 2 + 0 \end{bmatrix} \\
 &= \begin{bmatrix} 3 \\ 6 \\ 5 \\ -1 \end{bmatrix} = \mathbf{b}
 \end{aligned}$$

Expressing the solution set as the general solution.

$$\mathcal{S} = \left\{ \mathbf{x} \in \mathbb{R}^5 : \mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} -1 \\ -2 \\ 0 \\ -1 \\ -1 \end{bmatrix}, \lambda \in \mathbb{R} \right\}$$

We verified this for $\lambda = 0$. Let us do it again for $\lambda = 1$, which gives $\mathbf{x}_2 = [2, -2, 0, -2, -1]^T$. The following calculations verify that \mathbf{Ax}_2 also produces \mathbf{b} .

$$\begin{aligned}
 \mathbf{Ax}_2 &= \begin{bmatrix} 1 & -1 & 0 & 0 & 1 \\ 1 & 1 & 0 & -3 & 0 \\ 2 & -1 & 0 & 1 & -1 \\ -1 & 2 & 0 & -2 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ -2 \\ 0 \\ -2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1(2) + (-1)(-2) + 0(0) + 0(-2) + 1(-1) \\ 1(2) + 1(-2) + 0(0) + (-3)(-2) + 0(-1) \\ 2(2) + (-1)(-2) + 0(0) + 1(-2) + (-1)(-1) \\ (-1)(2) + 2(-2) + 0(0) + (-2)(-2) + (-1)(-1) \end{bmatrix} \\
 &= \begin{bmatrix} 2 + 2 + 0 + 0 - 1 \\ 2 - 2 + 0 + 6 + 0 \\ 4 + 2 + 0 - 2 + 1 \\ -2 - 4 + 0 + 4 + 1 \end{bmatrix} \\
 &= \begin{bmatrix} 3 \\ 6 \\ 5 \\ -1 \end{bmatrix} = \mathbf{b}
 \end{aligned}$$

2.6 Finding the solution to the following inhomogeneous linear equation. Start by finding the reduced echelon form.

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

Both x_1 and x_3 are irrelevant, and x_6 is a free variable. The general solution can be expressed as:

$$\mathcal{S} = \left\{ \mathbf{x} \in \mathbb{R}^6 : \mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \lambda \in \mathbb{R} \right\}$$

Verifying the solution for $\lambda = 2$. This gives us:

$$\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \\ -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \\ -1 \\ -2 \end{bmatrix}$$

By multiplying this vector to \mathbf{A} we can confirm that we get the solution \mathbf{b} .

$$\begin{aligned} \mathbf{Ax}_2 &= \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 0(0) + 1(3) + 0(0) + 0(0) + 1(-1) + 0(-2) \\ 0(0) + 0(3) + 0(0) + 1(0) + 1(-1) + 0(-2) \\ 0(0) + 1(3) + 0(0) + 0(0) + 0(-1) + 1(-2) \end{bmatrix} \\ &= \begin{bmatrix} 0 + 3 + 0 + 0 - 1 + 0 \\ 0 + 0 + 0 + 0 - 1 + 0 \\ 0 + 3 + 0 + 0 + 0 - 2 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \mathbf{b} \end{aligned}$$

2.7

We are solving $\mathbf{Ax} = \mathbf{12x}$ under the condition that $x_1 + x_2 + x_3 = 1$. Writing out the equations:

$$\begin{aligned}6x_1 + 4x_2 + 3x_3 &= 12x_1 \\6x_1 + 9x_3 &= 12x_2 \\8x_2 &= 12x_3\end{aligned}$$

By subtracting the right hand sides, we get the following, equivalent system of linear equations.

$$\begin{aligned}-6x_1 + 4x_2 + 3x_3 &= 0 \\6x_1 - 12x_2 + 9x_3 &= 0 \\8x_2 - 12x_3 &= 0\end{aligned}$$

Using this we get the this augmented matrix and find the row reduced echelon form.

$$[(\mathbf{A} - 12I_3)|\mathbf{0}] = \left[\begin{array}{ccc|c} -6 & 4 & 3 & 0 \\ 6 & -12 & 9 & 0 \\ 0 & 8 & -12 & 0 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -\frac{3}{2} & 0 \\ 0 & 1 & -\frac{3}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

So x_3 is a free variable, and x_1 and x_2 are expressed in terms of x_3 . This is where the extra condition $x_1 + x_2 + x_3 = 1$ comes into play. Writing the reduced echelon form as equations, we have:

$$\begin{aligned}x_1 - \frac{3}{2}x_3 &= 0 & \implies & x_1 = \frac{3}{2}x_3 \\x_2 - \frac{3}{2}x_3 &= 0 & \implies & x_2 = \frac{3}{2}x_3\end{aligned}$$

Replacing x_1 and x_2 with our expressions.

$$x_1 + x_2 + x_3 = 1 \implies \frac{3}{2}x_3 + \frac{3}{2}x_3 + x_3 = 1 \implies 4x_3 = 1 \implies x_3 = \frac{1}{4}$$

With this we can find the values for x_1 and x_2 .

$$x_1 = \frac{3}{2}x_3 = \frac{3}{2} \left(\frac{1}{4} \right) = \frac{3}{8} \quad \text{and} \quad x_2 = \frac{3}{2}x_3 = \frac{3}{2} \left(\frac{1}{4} \right) = \frac{3}{8}$$

The solution is $\mathbf{x} = [\frac{3}{8}, \frac{3}{8}, \frac{1}{4}]^T$. To verify the solution, we see that $12\mathbf{x} = [4.5, 4.5, 3]^T$. By multiplying \mathbf{Ax} we can check if we get this result. Without showing all the calculations:

$$\begin{bmatrix} 6 & 4 & 3 \\ 6 & 0 & 9 \\ 0 & 8 & 0 \end{bmatrix} \begin{bmatrix} 3/8 \\ 3/8 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 4.5 \\ 4.5 \\ 3 \end{bmatrix}$$

2.8 Finding inverse of matrices. We augment a matrix with the identity matrix and try to find the row reduced echelon form. If this is not possible, the matrix is not invertible.

a.

$$[\mathbf{A}|I_3] = \left[\begin{array}{ccc|ccc} 2 & 3 & 4 & 1 & 0 & 0 \\ 3 & 4 & 5 & 0 & 1 & 0 \\ 4 & 5 & 6 & 0 & 0 & 1 \end{array} \right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 0 & -5 & 4 \\ 0 & 1 & 2 & 0 & 4 & -3 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{array} \right]$$

There is a linear dependence between the third column and the two first columns of the matrix, hence it is not invertible.

b.

$$[\mathbf{A}|I_4] = \left[\begin{array}{cccc|cccc} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{array} \right] \sim \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & -2 \end{array} \right] = [I_4|\mathbf{A}^{-1}]$$

In this case we were able to find the row-reduced echelon form, which means the matrix is invertible. By multiplying $\mathbf{A}\mathbf{A}^{-1}$ we should get the identity matrix I_4 . Calculations to verify:

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

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

$$= \begin{bmatrix} 0+0+1+0 & -1+0+1+0 & 0+0+0+0 & 1+0-1+0 \\ 0-1+1+0 & 0+0+1+0 & 0+0+0+0 & 0+1-1+0 \\ 0-1+0+1 & -1+0+0+1 & 0+0+0+1 & 1+1+0-2 \\ 0-1+1+0 & -1+0+1+0 & 0+0+0+0 & 1+1-1+0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = I_4$$

Phew! The result is the identity matrix I_4 .

2.9 Determining whether four proposed sets are subspaces of \mathbb{R}^3 . There are two properties that needs to be checked.

1. The set contains the null vector $\mathbf{0} = [0, 0, 0]^T$.
2. The set is closed under scaling (multiplying with a constant) and addition.

A subset is not a subspace if we can find a single counterexample to any of these properties.

a.

The set A is defined as

$$A = \{\lambda, \lambda + \mu^3, \lambda - \mu^3 \mid \lambda, \mu \in \mathbb{R}\}$$

Showing that $\mathbf{0} \in A$ is easy as it is true when $\lambda = 0$ and $\mu = 0$.

Assuming some constant $c \in \mathbb{R}$ and $\mathbf{x} \in A$ with $a, b \in \mathbb{R}$ such that $\mathbf{x} = [a, a + b^3, a - b^3]$. We want to check if $c\mathbf{x} \in A$. Let's also define $q := c^{\frac{1}{3}}$ (if c is negative, this is the real root).

$$c\mathbf{x} = c \begin{bmatrix} a \\ a + b^3 \\ a - b^3 \end{bmatrix} = \begin{bmatrix} ca \\ ca + cb^3 \\ ca - cb^3 \end{bmatrix} = \begin{bmatrix} ca \\ ca + (qb)^3 \\ ca - (qb)^3 \end{bmatrix} = \begin{bmatrix} \lambda \\ \lambda + \mu^3 \\ \lambda - \mu^3 \end{bmatrix}$$

when $\lambda := ca$ and $\mu := qb$. This shows that A is closed under scaling. Next we assume we have to points in A , given by $\mathbf{x} = [a, a + b^3, a - b^3]$ and $\mathbf{y} = [x, x + y^3, x - y^3]$ for $a, b, x, y \in \mathbb{R}$. Adding them together.

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} a \\ a + b^3 \\ a - b^3 \end{bmatrix} + \begin{bmatrix} x \\ x + y^3 \\ x - y^3 \end{bmatrix} = \begin{bmatrix} (a + x) \\ (a + x) + (b^3 + y^3) \\ (a + x) - (b^3 + y^3) \end{bmatrix} = \begin{bmatrix} \lambda \\ \lambda + \mu^3 \\ \lambda - \mu^3 \end{bmatrix}$$

when $\lambda := (a + x)$ and $\mu^3 := (b^3 + y^3)$. With the properties verified, we can conclude that A is a subspace.

b.

Checking this set:

$$B = \{\lambda^2, -\lambda^2, 0 \mid \lambda \in \mathbb{R}\}$$

When $\lambda = 0$ we have $\mathbf{0} \in B$.

But B is not closed under scaling. A counterexample is when $\mathbf{z} \in B$ with $\lambda = 2$ and $c = -1$. In this case:

$$c\mathbf{z} = (-1) \begin{bmatrix} (2)^2 \\ -(2)^2 \\ 0 \end{bmatrix} = (-1) \begin{bmatrix} 4 \\ -4 \\ 0 \end{bmatrix} = \begin{bmatrix} -4 \\ 4 \\ 0 \end{bmatrix}$$

There is no $\lambda \in \mathbb{R}$ we can choose such that $\lambda^2 = -4$ or $-\lambda^2 = 4$. We would need complex numbers which violates the condition that $\lambda \in \mathbb{R}$. Hence B is NOT a subspace.

c.

We have some fixed value $\gamma \in \mathbb{R}$. The set is defined as:

$$C = \{\xi_1, \xi_2, \xi_3 \mid \xi_1 - 2\xi_2 + 3\xi_3 = \gamma\}$$

If $\gamma \neq 0$, then $\mathbf{0} \notin C$ which means this can only be a subspace when $\gamma = 0$.

Assuming that $\gamma = 0$ we can check if it satisfies closure under scaling/addition. Assuming $\mathbf{x} \in C$ where $\mathbf{x} = [a_1, a_2, a_3]$ and $a_1 - 2a_2 + 3a_3 = 0$ and $c \in \mathbb{R}$ and $c \neq 0$.

$$c\mathbf{x} = \begin{bmatrix} ca_1 \\ ca_2 \\ ca_3 \end{bmatrix}$$

Checking the condition for membership in C .

$$a_1 - 2a_2 + 3a_3 = 0 \Rightarrow c(a_1 - 2a_2 + 3a_3) = c(0) \Rightarrow ca_1 - 2ca_2 + 3ca_3 = 0$$

This shows the condition is met, so C is closed under scaling.

Introducing $\mathbf{y} = [b_1, b_2, b_3]$ and assuming $b_1 - 2b_2 + 3b_3 = 0$ we can add \mathbf{x} with \mathbf{y} and check if it's still a member of C .

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{bmatrix}$$

We need to check if this satisfies the condition for membership in C .

$$\begin{aligned} (a_1 + b_1) - 2(a_2 + b_2) + 3(a_3 + b_3) &= \\ a_1 + b_1 - 2a_2 - 2b_2 + 3a_3 + 3b_3 &= \\ (a_1 - 2a_2 + 3a_3) + (b_1 - 2b_2 + 3b_3) &= 0 + 0 = 0 \end{aligned}$$

(In the above step we used the assumptions we made earlier on a_i and b_i). This means:

$$(a_1 + b_1) - 2(a_2 + b_2) + 3(a_3 + b_3) = 0$$

This shows that C is closed under addition which means that C is a subspace if $\gamma = 0$.

d.

The following set:

$$D = \{(\xi_1, \xi_2, \xi_3) \mid \xi_2 \in \mathbb{Z}\}$$

is NOT a subspace since it is not closed under scaling. Counterexample: take the element $\mathbf{y} = [2, 1, 2]^T \in D$ and scale with $c = 1.1$. Then $1.1\mathbf{y} = [2.2, 1.1, 2.2]^T$ and since $1.1 \notin \mathbb{Z}$ it is not closed under scaling and therefore NOT a subspace.

2.10. Finding out if the following vectors are linearly independent.

a.

We collect the terms in a matrix and find the reduced echelon form to answer the question.

$$[\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3] = \begin{bmatrix} 2 & 1 & 3 \\ -1 & 1 & -3 \\ 3 & -2 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

As we can see, the third column depends on the first two: $\mathbf{x}_3 = 2\mathbf{x}_1 - \mathbf{x}_2$. Verifying:

$$2\mathbf{x}_1 - \mathbf{x}_2 = 2 \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 6 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 4-1 \\ -2-1 \\ 6+2 \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \\ 8 \end{bmatrix} = \mathbf{x}_3$$

b.

Same procedure.

$$[\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3] = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

In this case the rows are linearly independent (not surprising since there are five rows).

2.11. We write out the augmented matrix $[\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3 | \mathbf{y}]$ and find the row reduced echelon form to see how \mathbf{y} can be written as a linear combination of the other columns.

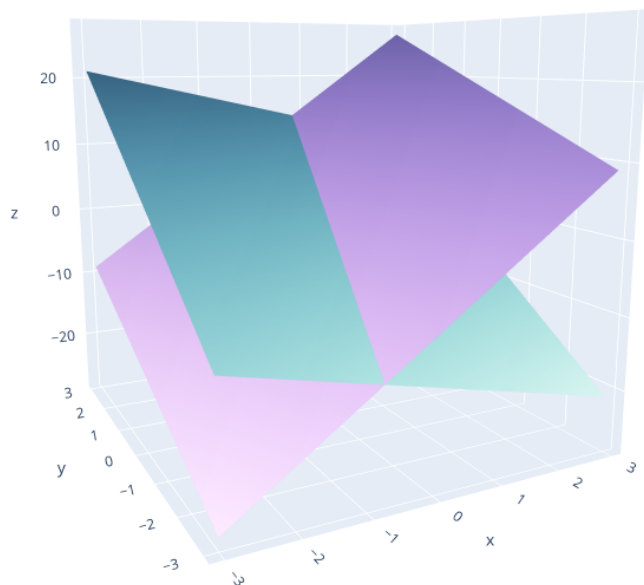
$$[\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3 | \mathbf{y}] = \left[\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 1 & 2 & -1 & -2 \\ 1 & 3 & 1 & 5 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & -6 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 2 \end{array} \right]$$

So, $\mathbf{y} = -6\mathbf{x}_1 + 3\mathbf{x}_2 + 2\mathbf{x}_3$. Verifying:

$$\begin{aligned} -6\mathbf{x}_1 + 3\mathbf{x}_2 + 2\mathbf{x}_3 &= -6 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -6 \\ -6 \\ -6 \end{bmatrix} + \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} + \begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} -6+3+4 \\ -6+6-2 \\ -6+9+2 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix} = \mathbf{y} \end{aligned}$$

2.12. Finding a basis of $U_1 \cap U_2$ for two subspaces $U_1, U_2 \subset \mathbb{R}^4$.

Let's first consider a simplified example where U_1 and U_2 are subspaces of \mathbb{R}^3 and that they are both planes running through the origin. Their intersection would be a line going through the origin. Here is an example (code enclosed in the **Appendix**):



In this example we have the following surfaces:

$$(I) \quad z_1 = 6x + 3y$$
$$(II) \quad z_2 = -2x + 5y$$

We can find a parametric line that describes the intersection with some algebra. Since we have two equations and three variables, we set $t := x$ and solve the equation for y and z in terms of t .

$$(II) \quad z = -2t + 5y \Rightarrow 2t = 5y - z \stackrel{(I)}{\Rightarrow} 2t = 5y - 6t - 3y \Rightarrow 8t = 2y \Rightarrow y = 4t$$

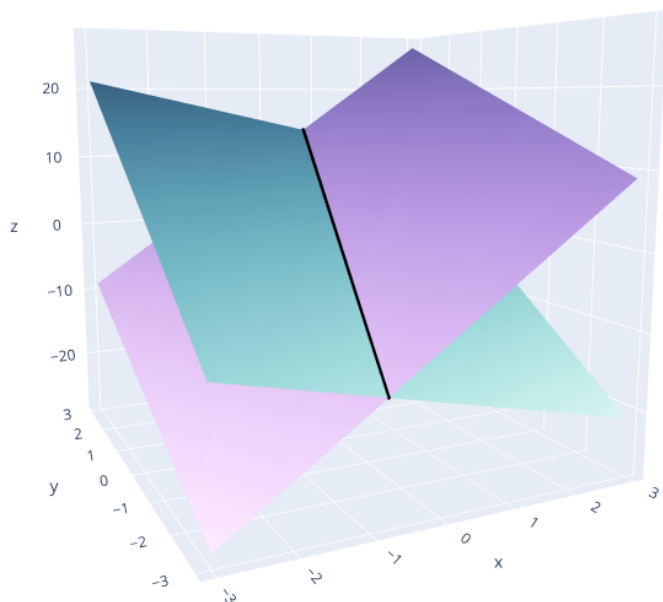
Replacing the value for y into (I):

$$(I) \quad z = 6t + 3y \Rightarrow z = 6t + 3(4t) \Rightarrow z = 6t + 12t \Rightarrow z = 18t$$

So, expressed in terms of t , the the parametric line is given as:

$$\begin{bmatrix} t \\ 4t \\ 18t \end{bmatrix} = t \begin{bmatrix} 1 \\ 4 \\ 18 \end{bmatrix}$$

Here is a plot showing that this parametric line is indeed the intersection between the planes. (Code enclosed in **Appendix**).



In other words, the vector we found spans this exact line. Let us solve it using just the bases which is the intention of this exercise. So how do we determine the span from a plane? Any linearly independent vectors in a subspace will be a span, and for our planes we just need to find two vectors (because they are 2D planes). In the case of equation (I), we can simply input $x = 2, y = 1$ and $x = 1, y = 2$ to calculate z and find two independent vectors.

$$z = 6(2) + 3(1) = 12 + 3 = 15, \quad z = 6(1) + 3(2) = 6 + 6 = 12$$

We can do the same for (II).

$$z = -2(2) + 5(1) = -4 + 5 = 1, \quad z = -2(1) + 5(2) = -2 + 10 = 8$$

This gives us the span for each plane.

$$U_1 = \text{span}\{\mathbf{u}_1, \mathbf{u}_2\} = \text{span}\left\{\begin{bmatrix} 2 \\ 1 \\ 15 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 12 \end{bmatrix}\right\}, \quad U_2 = \text{span}\{\mathbf{w}_1, \mathbf{w}_2\} = \text{span}\left\{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 8 \end{bmatrix}\right\}$$

Now to find the span of $U_1 \cap U_2$. Assume that we have some $\mathbf{v} \in U_1 \cap U_2$. By properties of a basis, we can write this as $\mathbf{v} = a\mathbf{u}_1 + b\mathbf{u}_2$ and $\mathbf{v} = c\mathbf{w}_1 + d\mathbf{w}_2$ for some $a, b, c, d \in \mathbb{R}$. Since $\mathbf{v} - \mathbf{v} = \mathbf{0}$, we get the following equation:

$$\mathbf{v} - \mathbf{v} = a\mathbf{u}_1 + b\mathbf{u}_2 - c\mathbf{w}_1 - d\mathbf{w}_2 = \mathbf{0}$$

Writing this out as an augmented matrix and finding the row reduced echelon form, we can find the unknown values a, b, c, d .

$$\left[\begin{array}{cccc|c} 2 & 1 & -2 & -1 & 0 \\ 1 & 2 & -1 & -2 & 0 \\ 15 & 12 & -1 & -8 & 0 \end{array} \right] \sim \left[\begin{array}{cccc|c} 1 & 0 & 0 & \frac{2}{7} & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & \frac{2}{7} & 0 \end{array} \right]$$

Important note. Since U_1 and U_2 are 2-dimensional planes, $\dim(U_1) = 2$ and $\dim(U_2) = 2$. The last matrix has rank 3, so $\dim(U_1 + U_2) = 3$. From the following formula we can calculate $\dim(U_1 \cap U_2)$:

$$\dim(U_1 \cap U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 + U_2) = 2 + 2 - 3 = 1.$$

This verifies that the intersection is spanned by 1 vector, i.e. it is a line.

Above, d is a free variable. Writing the solution as a linear equation.

$$\begin{aligned} a &= -\frac{2}{7}d \\ b &= d \\ c &= -\frac{2}{7}d \end{aligned}$$

We can make the special choice $d = \frac{7}{3}$, which gives us:

$$\begin{aligned} a &= -\frac{2}{7} \left(\frac{7}{3} \right) = -\frac{2}{3} \\ b &= \frac{7}{3} \\ c &= -\frac{2}{7} \left(\frac{7}{3} \right) = -\frac{2}{3} \end{aligned}$$

Now that we know a, b, c, d we can calculate \mathbf{v} .

$$\mathbf{v} = a\mathbf{u}_1 + b\mathbf{u}_2 = -\frac{2}{3} \begin{bmatrix} 2 \\ 1 \\ 15 \end{bmatrix} + \frac{7}{3} \begin{bmatrix} 1 \\ 2 \\ 12 \end{bmatrix} = \begin{bmatrix} -\frac{4}{3} \\ -\frac{2}{3} \\ -10 \end{bmatrix} + \begin{bmatrix} \frac{7}{3} \\ \frac{14}{3} \\ 28 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 18 \end{bmatrix}$$

Not necessary, but we can also calculate \mathbf{v} from U_2 .

$$\mathbf{v} = c\mathbf{w}_1 + d\mathbf{w}_2 = -\frac{2}{3} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \frac{7}{3} \begin{bmatrix} 1 \\ 2 \\ 8 \end{bmatrix} = \begin{bmatrix} -\frac{4}{3} \\ -\frac{2}{3} \\ -\frac{2}{3} \end{bmatrix} + \begin{bmatrix} \frac{7}{3} \\ \frac{14}{3} \\ \frac{56}{3} \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 18 \end{bmatrix}$$

This corresponds to the parametrized line that we found earlier. We have now shown that this is the basis of $U_1 \cap U_2$.

$$U_1 \cap U_2 = \text{span} \left\{ \begin{bmatrix} 1 \\ 4 \\ 18 \end{bmatrix} \right\}$$

(We selected $d = \frac{7}{3}$ to get the same basis vector that we found earlier. Any other value for d would simply have given a scaled version of the same vector, which would ultimately make the same line as in the plot).

So, now to the actual exercise! We have subspaces $U_1 \subset \mathbb{R}^4$ and $U_2 \subset \mathbb{R}^4$ and the spans are given by:

$$U_1 = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \right\}, \quad U_2 = \text{span} \left\{ \begin{bmatrix} -1 \\ -2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 6 \\ -2 \\ -1 \end{bmatrix} \right\}$$

which we will simplify as $U_1 = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and $U_2 = \text{span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$. We will determine a basis for $U_1 \cap U_2$. As in the simplified example above, we assume there is some point $\mathbf{v} \in U_1 \cap U_2$, so for $a, b, c, d, e, f \in \mathbb{R}$ we can write:

$$\mathbf{v} = a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_3$$

and

$$\mathbf{v} = d\mathbf{w}_1 + e\mathbf{w}_2 + f\mathbf{w}_3$$

By taking $\mathbf{v} - \mathbf{v}$ we get a homogeneous equation which we can express as the augmented matrix from which we get the reduced echelon form.

$$\left[\begin{array}{cccccc|c} 1 & 2 & -1 & 1 & -2 & 3 & 0 \\ 1 & -1 & 1 & 2 & 2 & -6 & 0 \\ -3 & 0 & -1 & -2 & 0 & 2 & 0 \\ 1 & -1 & 1 & -1 & 0 & 1 & 0 \end{array} \right] \sim \left[\begin{array}{cccccc|c} 1 & 0 & \frac{1}{3} & 0 & -\frac{4}{9} & \frac{8}{9} & 0 \\ 0 & 1 & -\frac{2}{3} & 0 & -\frac{10}{9} & \frac{20}{9} & 0 \\ 0 & 0 & 0 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \quad (*)$$

The free variables are c, d and f . The equations are:

$$a = -\frac{1}{3}c + \frac{4}{9}e - \frac{8}{9}f$$

$$b = \frac{2}{3}c + \frac{10}{9}e - \frac{20}{9}f$$

$$c = c$$

$$d = -\frac{2}{3}e + \frac{7}{3}f$$

$$e = e$$

$$f = f$$

Setting $c = e = f = 9$, we get:

$$a = -\frac{1}{3}(9) + \frac{4}{9}(9) - \frac{8}{9}(9) = -3 + 4 - 8 = -7$$

$$b = \frac{2}{3}(9) + \frac{10}{9}(9) - \frac{20}{9}(9) = 6 + 10 - 20 = -4$$

$$c = 9$$

$$d = -\frac{2}{3}(9) + \frac{7}{3}(9) = -6 + 21 = 15$$

$$e = 9$$

$$f = 9$$

Now we can calculate \mathbf{v} .

$$\begin{aligned}\mathbf{v} = a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_3 &= (-7) \begin{bmatrix} 1 \\ 1 \\ -3 \\ 1 \end{bmatrix} + (-4) \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \end{bmatrix} + 9 \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -7 \\ -7 \\ 21 \\ -7 \end{bmatrix} + \begin{bmatrix} -8 \\ 4 \\ 0 \\ 4 \end{bmatrix} + \begin{bmatrix} -9 \\ 9 \\ -9 \\ 9 \end{bmatrix} \\ &= \begin{bmatrix} -7 - 8 - 9 \\ -7 + 4 + 9 \\ 21 + 0 - 9 \\ -7 + 4 + 9 \end{bmatrix} \\ &= \begin{bmatrix} -24 \\ 6 \\ 12 \\ 6 \end{bmatrix} \\ &= 6 \begin{bmatrix} -4 \\ 1 \\ 2 \\ 1 \end{bmatrix}\end{aligned}$$

According to these calculations,

$$U_1 \cap U_2 = \text{span} \left\{ \begin{bmatrix} -4 \\ 1 \\ 2 \\ 1 \end{bmatrix} \right\}$$

Not showing the calculations, but as we should have checked earlier: $\dim(U_1) = 2$ and $\dim(U_2) = 2$. From (*) above we see that $\dim(U_1) + \dim(U_2) = 3$ since the augmented matrix has rank 3.

$$\dim(U_1 \cap U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 + U_2) = 2 + 2 - 3 = 1.$$

This shows that the intersection is a line. (Would have simplified the calculations a bit, but whaddya gonna do?).

2.13. The subspace U_i is the solution to the homogeneous equation systems $\mathbf{A}_i \mathbf{x} = \mathbf{0}$ for $i = 1, 2$ where

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

a. and b.

To find the solution to the null space (or kernel) we will use the "Minus-1" trick from section **2.3.3**. First we find the row reduced echelon form.

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -2 & -1 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This is easily verified since the third column is just the sum of the first two columns. By the 'Minus-1' trick, we extend the matrix into the augmented, square matrix $\tilde{\mathbf{A}}_1$ and we insert a '-1' beneath the two 1s, marked in bold text and on the last column, so the diagonal contains either 1 or -1.

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

All the rows with -1 on the diagonal will be the span of the solution space U_1 . (Both vectors in the span will give $\mathbf{0}$ when multiplied with \mathbf{A}_1 .) This is also a basis for U_1 and we can see that $\dim(U_1) = 2$.

$$\text{span}(U_1) = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix} \right\}$$

We repeat the same procedure for \mathbf{A}_2 .

$$\mathbf{A}_2 = \begin{bmatrix} 3 & -3 & 0 \\ 1 & 2 & 3 \\ 7 & -5 & 2 \\ 3 & -1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Again, the third column is the sum of the first two columns which is easily seen (at least in retrospect!). Now, we have the exact same row reduced echelon form as earlier, and the "Minus-1" trick will yield the same exact span/basis. We can note that $\dim(U_2) = 2$.

$$\text{span}(U_2) = \left\{ \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix} \right\}$$

c.

The null spaces for \mathbf{A}_1 and \mathbf{A}_2 are spanned by the same vectors. In other words they are completely overlapping planes. Hence, the basis for $U_1 \cap U_2$ is the same as for U_1 and U_2 .

2.14.

a.

To find the dimensions we count the pivot columns in $\text{rref}(A_i)$. The matrices here are the same ones from the previous exercise.

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -2 & -1 \\ 2 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

There are two pivot columns, so $\dim(U_1) = 2$.

$$\mathbf{A}_2 = \begin{bmatrix} 3 & -3 & 0 \\ 1 & 2 & 3 \\ 7 & -5 & 2 \\ 3 & -1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

There are two pivot columns, so $\dim(U_2) = 2$.

b.

To determine the basis we simply have to find the pivot columns, which we saw are the first two columns. The basis of U_1 is:

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} \right\} = \{\mathbf{u}_1, \mathbf{u}_2\}$$

It's exactly the same for U_2 . Its basis is:

$$\left\{ \begin{bmatrix} 3 \\ 1 \\ 7 \\ 3 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \\ -5 \\ -1 \end{bmatrix} \right\} = \{\mathbf{w}_1, \mathbf{w}_2\}$$

c.

First we calculate the dimension of $U_1 \cap U_2$. We need to find the dimension of $U_1 + U_2$ first.

$$[\mathbf{A}_1 \ \mathbf{A}_2] = \begin{bmatrix} 1 & 0 & 1 & 3 & -3 & 0 \\ 1 & -2 & -1 & 1 & 2 & 3 \\ 2 & 1 & 3 & 7 & -5 & 2 \\ 1 & 0 & 1 & 3 & -1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 & 3 & -3 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

There are three pivot columns so $\dim(U_1 + U_2) = 3$. We can now calculate the dimension of $U_1 \cap U_2$.

$$\dim(U_1 \cap U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 + U_2) = 2 + 2 - 3 = 1$$

Solving this the same way as earlier. Assuming $\mathbf{v} \in U_1 \cap U_2$, then, for $a, b, c, d \in \mathbb{R}$ we have

$$\begin{aligned} \mathbf{v} &= a\mathbf{u}_1 + b\mathbf{u}_2 \\ \mathbf{v} &= c\mathbf{w}_1 + d\mathbf{w}_2 \end{aligned}$$

From $\mathbf{v} - \mathbf{v}$ we get $a\mathbf{u}_1 + b\mathbf{u}_2 - c\mathbf{w}_1 - d\mathbf{w}_2 = \mathbf{0}$ and can set up the following matrix in order to solve a, b, c, d . (NOTE: These are supposed to be augmented matrices).

$$\left[\begin{array}{cccc|c} 1 & 0 & -3 & 3 & 0 \\ 1 & -2 & -1 & -2 & 0 \\ 2 & 1 & -7 & 5 & 0 \\ 1 & 0 & -3 & 1 & 0 \end{array} \right] \sim \left[\begin{array}{cccc|c} 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

Writing the equation, and setting the free variable $c = 1$.

$$\begin{array}{rcl} a - 3c = 0 & & a = 3 \\ b - c = 0 & \implies & b = 1 \\ c = c & & c = 1 \\ d = 0 & & d = 0 \end{array}$$

Since $d = 0$, we get from U_2 :

$$\mathbf{v} = c\mathbf{w}_1 + d\mathbf{w}_2 = \mathbf{w}_1 + \mathbf{0} = \begin{bmatrix} 3 \\ 1 \\ 7 \\ 3 \end{bmatrix}$$

We can also verify with calculating from U_1 .

$$\mathbf{v} = a\mathbf{u}_1 + b\mathbf{u}_2 = 3\mathbf{u}_1 + \mathbf{u}_2 = 3 \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 6 \\ 3 \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 7 \\ 3 \end{bmatrix}$$

The basis for $U_1 \cap U_2$ is therefore:

$$\left\{ \begin{bmatrix} 3 \\ 1 \\ 7 \\ 3 \end{bmatrix} \right\}$$

2.15. We have the following sets:

$$F = \{(x, y, z) \in \mathbb{R}^3 \mid x + y - z = 0\}$$

$$G = \{(a - b, a + b, a - 3b) \in \mathbb{R}^3 \mid a, b \in \mathbb{R}\}$$

a.

Showing that F is a subspace of \mathbb{R}^3 . It is obvious that $\mathbf{0} \in F$, so we just need to verify that it is closed under scaling and addition. Assume $c \in \mathbb{R}$ and $(x, y, z) \in F$ which means $x + y - z = 0$. Then:

$$cx + cy - cz = c(x + y - z) = c(0) = 0 \Rightarrow cx + cy + cz = 0$$

so $(cx, cy, cz) \in F$. Next we assume $(x_1, y_1, z_1), (x_2, y_2, z_2) \in F$ so $x_i + y_i - z_i = 0$ for $i = 1, 2$. When adding them we get $(x_1 + x_2, y_1 + y_2, z_1 + z_2)$. Verifying that this is in F :

$$(x_1 + x_2) + (y_1 + y_2) - (z_1 + z_2) = (x_1 + y_1 - z_1) + (x_2 + y_2 - z_2) = 0 + 0 = 0$$

Hence,

$$(x_1 + x_2) + (y_1 + y_2) - (z_1 + z_2) = 0$$

which shows F is closed under addition and therefore a subspace of \mathbb{R}^3 .

Next we show that G is a subspace. By selecting $a = 0$ and $b = 0$, we get $(0, 0, 0)$, so $\mathbf{0} \in G$. Next we show scaling. For some $c \in \mathbb{R}$, and $(a' - b', a' + b', a' - 3b') \in G$, the scaled vector is $(ca' - cb', ca' + cb', ca' - 3cb')$. We define $a := ca' \in \mathbb{R}$ and $b := cb' \in \mathbb{R}$ so we have $(a - b, a + b, a - 3b)$ which is obviously in G . Finally we show that G is closed under addition. Assuming $(a_1 - b_1, a_1 + b_1, a_1 - 3b_1) \in G$ and $(a_2 - b_2, a_2 + b_2, a_2 - 3b_2) \in G$. By adding them:

$$\begin{bmatrix} a_1 - b_1 \\ a_1 + b_1 \\ a_1 - 3b_1 \end{bmatrix} + \begin{bmatrix} a_2 - b_2 \\ a_2 + b_2 \\ a_2 - 3b_2 \end{bmatrix} = \begin{bmatrix} a_1 - b_1 + a_2 - b_2 \\ a_1 + b_1 + a_2 + b_2 \\ a_1 - 3b_1 + a_2 - 3b_2 \end{bmatrix} = \begin{bmatrix} (a_1 + a_2) - (b_1 + b_2) \\ (a_1 + a_2) + (b_1 + b_2) \\ (a_1 + a_2) - 3(b_1 + b_2) \end{bmatrix} = \begin{bmatrix} a_3 - b_3 \\ a_3 + b_3 \\ a_3 - 3b_3 \end{bmatrix}$$

where we defined $a_3 := a_1 + a_2$ and $b_3 := b_1 + b_2$ which are real numbers. This shows G is closed under addition and that G is a subspace.

b.

We will find the intersection $F \cap G$. This set is defined as the set where both conditions are true:

$$F \cap G = \left\{ \begin{bmatrix} a - b \\ a + b \\ a - 3b \end{bmatrix} \in \mathbb{R}^3 \mid \begin{matrix} (a - b) & + & (a + b) & - & (a - 3b) & = & 0 \\ & x & y & & z & & \end{matrix} \text{ and } a, b \in \mathbb{R} \right\}$$

(Marking the new conditions that are x, y, z in F). Simplifying the expression:

$$(a - b) + (a + b) - (a - 3b) = 0 \Rightarrow a + 3b = 0 \Rightarrow a = -3b$$

Replacing $a = -3b$ into the vectors:

$$\begin{bmatrix} a - b \\ a + b \\ a - 3b \end{bmatrix} = \begin{bmatrix} -3b - b \\ -3b + b \\ -3b - 3b \end{bmatrix} = \begin{bmatrix} -4b \\ -2b \\ -6b \end{bmatrix} = -2b \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = w \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

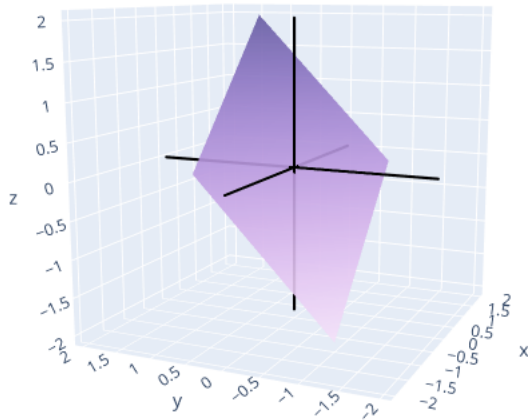
Since $b \in \mathbb{R}$ is any real number, we can use $w \in \mathbb{R}$ defined as $w := -2b$ instead to get a scaled down version. So, in conclusion:

$$F \cap G = \left\{ w \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \in \mathbb{R}^3 \mid w \in \mathbb{R} \right\}$$

As a verification we can see that $[2, 1, 3]^T \in F$ since $2 + 1 - 3 = 0$ and $[2, 1, 3]^T \in G$ when $a = 3/2$ and $b = -1/2$ which shows that both conditions are satisfied.

c.

Finding a basis for F . The condition is that $x + y - z = 0$ which means $z = x + y$ which is easily plottable as a surface (code example in **Appendix**), with axes included:



To find a basis we just need any linearly independent vectors that span the set. In this case we have a 2D-plane so we just need two vectors. If we set in $x = 2, y = 1$ and then $x = 1, y = 2$ we get $z = 3$, but we also get a basis.

$$\left\{ \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\} = \{\mathbf{u}_1, \mathbf{u}_2\}$$

To find a basis for G note that for $a, b \in \mathbb{R}$ we have the following set of equalities:

$$\begin{bmatrix} a - b \\ a + b \\ a - 3b \end{bmatrix} = \begin{bmatrix} a \\ a \\ a \end{bmatrix} + \begin{bmatrix} -b \\ b \\ -3b \end{bmatrix} = a \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + b \begin{bmatrix} -1 \\ 1 \\ -3 \end{bmatrix}$$

so any member of G can be expressed as $a\mathbf{w}_1 + b\mathbf{w}_2$, but that is exactly the definition of a basis, so we can just read it off. This means the following is the basis for G , which is also a 2-dimensional plane:

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -3 \end{bmatrix} \right\} = \{\mathbf{w}_1, \mathbf{w}_2\}$$

Now to find a basis for $F \cap G$. First we make a few observations: $\dim(F) = 2$, $\dim(G) = 2$. To find the dimensions of $F + G$ we count the pivot columns of a matrix with the bases:

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

So, $\dim(F + G) = 3$. This means that $\dim(F \cap G) = 2 + 2 - 3 = 1$, i.e. it is a line in 3d-space.

Assuming $\mathbf{v} \in F \cap G$ we can write it as the following (assuming $x_1, x_2, x_3, x_4 \in \mathbb{R}$):

$$\begin{aligned} \mathbf{v} &= x_1 \mathbf{u}_1 + x_2 \mathbf{u}_2 \\ \mathbf{v} &= x_3 \mathbf{w}_1 + x_4 \mathbf{w}_2 \end{aligned}$$

And to find the parameters x_i we solve the equation: $x_1 \mathbf{u}_1 + x_2 \mathbf{u}_2 - x_3 \mathbf{w}_1 - x_4 \mathbf{w}_2 = \mathbf{0}$.

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

Written as equations, and setting the free variable $x_4 = -1/2$:

$$\begin{array}{rcl} x_1 + 2x_4 = 0 & x_1 = -2x_4 = -2\left(-\frac{1}{2}\right) & x_1 = 1 \\ x_2 = 0 & x_2 = 0 & x_2 = 0 \\ x_3 + 3x_4 = 0 & \implies x_3 = -3x_4 = -3\left(-\frac{1}{2}\right) \implies & x_3 = \frac{3}{2} \\ x_4 = x_4 & x_4 = -\frac{1}{2} & x_4 = -\frac{1}{2} \end{array}$$

Calculating \mathbf{v} from F :

$$\mathbf{v} = (1)\mathbf{u}_1 + (0)\mathbf{u}_2 = \mathbf{u}_1 + \mathbf{0} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

Calculating \mathbf{v} from G :

$$\mathbf{v} = x_3 \mathbf{w}_1 + x_4 \mathbf{w}_2 = \left(\frac{3}{2}\right) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \left(-\frac{1}{2}\right) \begin{bmatrix} -1 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \\ \frac{3}{2} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{3}{2} \end{bmatrix} = \begin{bmatrix} \frac{3}{2} + \frac{1}{2} \\ \frac{3}{2} - \frac{1}{2} \\ \frac{3}{2} + \frac{3}{2} \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

We get the same basis as we found in (b).

2.16. Recalling the definition of a linear mapping/homomorphism. A function $\Phi : V \rightarrow W$ is linear if for any $\mathbf{x}, \mathbf{y} \in V$ and any $a, b \in \mathbb{R}$:

$$\Phi(a\mathbf{x} + b\mathbf{y}) = a\Phi(\mathbf{x}) + b\Phi(\mathbf{y}).$$

a.

Checking if $\Phi : L^1([a, b]) \rightarrow \mathbb{R}$ is linear, when:

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

Yes, this is linear by standard properties of the integral. Assuming $f, g \in L^1([a, b])$ and $c_1, c_2 \in \mathbb{R}$. Then:

$$\Phi(c_1f + c_2g) = \int_a^b c_1 \cdot f(x) + c_2 \cdot g(x)dx = c_1 \int_a^b f(x)dx + c_2 \int_a^b g(x)dx = c_1\Phi(f) + c_2\Phi(g)$$

b.

Checking if $C^1 \rightarrow C^0$ is linear, when:

$$\Phi(f) = \frac{d}{dx}f = f'$$

Yes, this is linear by standard properties of the derivative. Assuming $f, g \in C^1$ and $c_1, c_2 \in \mathbb{R}$. Then:

$$\Phi(c_1f + c_2g) = \frac{d}{dx}(c_1f(x) + c_2g(x)) = c_1 \frac{d}{dx}f(x) + c_2 \frac{d}{dx}g(x) = c_1\Phi(f) + c_2\Phi(g)$$

c.

Checking the function $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ where $\Phi(x) = \cos(x)$. This is NOT a linear function.

Counterexample:

$$\cos(2 \cdot 2 + 2 \cdot 2) = \cos(4 + 4) = \cos(8) \approx -0.1455$$

$$2 \cos(2) + 2 \cos(2) = 4 \cos(2) \approx -1.6646$$

(It breaks both the properties... but that doesn't really matter. It is not a linear mapping).

d.

Checking the matrix operation $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ where

$$\Phi(\mathbf{x}) = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{A}\mathbf{x}.$$

Assuming $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$ and $c_1, c_2 \in \mathbb{R}$,

$$\Phi(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = \mathbf{A}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1\mathbf{A}\mathbf{x}_1 + c_2\mathbf{A}\mathbf{x}_2 = c_1\Phi(\mathbf{x}_1) + c_2\Phi(\mathbf{x}_2)$$

which shows that Φ is a linear mapping. (We used that matrices are distributive over other matrices/vectors. Property 2.19b from the text). Note that this actually proves that ALL matrices are linear mappings.

e.

Assuming $\theta \in [0, 2\pi)$ we want to check if the following rotation matrix is a linear mapping. For $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ given by

$$\Phi(\mathbf{x}) = \mathbf{A}\mathbf{x} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Assuming $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$ and $c_1, c_2 \in \mathbb{R}$,

$$\Phi(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = \mathbf{A}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1\mathbf{A}\mathbf{x}_1 + c_2\mathbf{A}\mathbf{x}_2 = c_1\Phi(\mathbf{x}_1) + c_2\Phi(\mathbf{x}_2)$$

This function is a linear mapping. The same argument can be used for all matrices.

2.17. We have a linear mapping $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ given by:

$$\Phi \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = \begin{bmatrix} 3x_1 + 2x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 - 3x_2 \\ 2x_1 + 3x_2 + x_3 \end{bmatrix}$$

- Finding the transformation matrix \mathbf{A}_Φ . We can simply read it from the matrix above.

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

This text book explains transformation matrices quite poorly, but to see that it is correct we can borrow an example from another Linear Algebra book. Here it states that the function $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ which for us corresponds to $\Phi(\mathbf{x}) = \mathbf{A}_\Phi\mathbf{x}$, and we can see the matrix directly from the results.

EXAMPLE 1 Let $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$, and

define a transformation $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ by $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$, so that

$$T(\mathbf{x}) = \mathbf{A}\mathbf{x} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 3x_2 \\ 3x_1 + 5x_2 \\ -x_1 + 7x_2 \end{bmatrix}$$

- The rank is determined by counting the pivots in $\text{rref}(\mathbf{A}_\Phi)$. This shows that $\text{rank}(\mathbf{A}_\Phi) = 3$.

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

- By the rank-nullity theorem we know that: $\text{rank}(\mathbf{A}_\Phi) = \dim(\text{Im}(\mathbf{A}_\Phi)) = 3$ and,

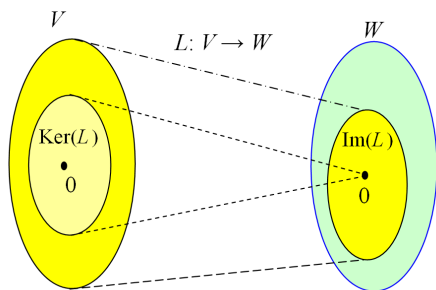
$$\dim(\mathbb{R}^3) = \text{rank}(\mathbf{A}_\Phi) + \dim(\ker(\mathbf{A}_\Phi)) = 3 + 0 \Rightarrow \dim(\ker(\mathbf{A}_\Phi)) = 0$$

(Using \mathbf{A}_Φ and Φ interchangeably, but they basically mean the same thing.)

2.18. A bunch of terms:

- Automorphism: the function $f : V \rightarrow V$ is *linear* and *bijective*.
- Linear: for any $\mathbf{x}, \mathbf{y} \in V$ and any $a, b \in \mathbb{R}$, $f(a\mathbf{x} + b\mathbf{y}) = af(\mathbf{x}) + bf(\mathbf{y})$.
- Bijective: f is *injective* and *surjective*. :)
- Injective: $\forall \mathbf{x}, \mathbf{y} \in V : f(\mathbf{x}) = f(\mathbf{y}) \Rightarrow \mathbf{x} = \mathbf{y}$
- Surjective: $f(V) = V$, i.e. the whole V domain is mapped to V and 'fills it out'.
- Define $id_V : V \rightarrow V$ as the identity mapping $id_V(\mathbf{x}) = \mathbf{x}$ for $\mathbf{x} \in V$.

This image can be helpful. (The yellow subset in W is V for this exercise).



Let E be a vector space. Let f and g be two automorphisms on E such that $f \circ g = id_E$ (i.e. $f \circ g$ is the identity mapping id_E). Show that $\ker(f) = \ker(g \circ f)$, $\text{Im}(g) = \text{Im}(g \circ f)$ and that $\ker(f) \cap \text{Im}(g) = \{\mathbf{0}_E\}$.

Claim: $\ker(f) = \ker(g \circ f)$

Proof.

\subset). Assume \mathbf{x} is some random point in $\ker(f)$, which means $f(\mathbf{x}) = \mathbf{0}_E$. Then $g \circ f(\mathbf{x}) = g(f(\mathbf{x})) = g(\mathbf{0}_E) = \mathbf{0}_E$ since the null point is always in the kernel, so $\mathbf{x} \in \ker(g \circ f)$, which proves $\ker(f) \subset \ker(g \circ f)$.

\supset). Now assume $\mathbf{x} \in \ker(g \circ f)$ so $g(f(\mathbf{x})) = \mathbf{0}_E$. Since the null point is always in the kernel, $g(\mathbf{0}_E) = \mathbf{0}_E$. By the injective property of g , $g(f(\mathbf{x})) = g(\mathbf{0}_E) \implies f(\mathbf{x}) = \mathbf{0}_E$ which means $\mathbf{x} \in \ker(f)$, so $\ker(g \circ f) \subset \ker(f)$.

By inclusion both ways, we have proved that $\ker(f) = \ker(g \circ f)$. ■

Claim: $\text{Im}(g) = \text{Im}(g \circ f)$

Proof.

Since g is surjective, $g(E) = E$. Since f is surjective, $f(E) = E$ and by the surjective property of f again we have $g(f(E)) = g(E) = E$. So, $\text{Im}(g) = E = \text{Im}(g \circ f)$. ■

Claim: $\ker(f) \cap \text{Im}(g) = \{\mathbf{0}_E\}$

Proof.

By the surjective property of g , $\text{Im}(g) = E$. Now assume some random $\mathbf{x} \in \ker(f)$, then $f(\mathbf{x}) = \mathbf{0}_E$. Since the null point is in the kernel, $f(\mathbf{0}_E) = \mathbf{0}_E$ which means $f(\mathbf{x}) = f(\mathbf{0}_E)$. By the injective property, $f(\mathbf{x}) = f(\mathbf{0}_E) \implies \mathbf{x} = \mathbf{0}_E$. From this we can conclude that $\ker(f) = \{\mathbf{0}_E\}$. This means $\ker(f) \cap \text{Im}(g) = \{\mathbf{0}_E\} \cap E = \{\mathbf{0}_E\}$. ■

2.19. We have an endomorphism (linear function which maps a domain back into itself) $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ with the transformation matrix with the corresponding RREF.

$$\mathbf{A}_\Phi = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

a.

All columns are pivot-columns, so the rank of the matrix is 3. The dimension of \mathbb{R}^3 is of course 3, which means the null space must have dimension 0, i.e. it is only a single point: the null point.

$$\ker(\Phi) = \{\mathbf{0}\}$$

Since there are three linearly independent columns, they span a 3d space, and since this is in \mathbb{R}^3 , they span the entire set.

$$\text{Im}(\Phi) = \mathbb{R}^3$$

b.

With the following ordered basis, we will determine the transformation matrix $\tilde{\mathbf{A}}_\Phi$.

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

Just some clarifications: if we have a point in \mathbb{R}^3 , the matrix will map it to some other point. As an example, consider the point $[2, 3, 1]^T$:

$$\begin{aligned} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} &= \begin{bmatrix} 1(2) + 1(3) + 0(1) \\ 1(2) + (-1)(3) + 0(1) \\ 1(2) + 1(3) + 1(1) \end{bmatrix} \\ &= \begin{bmatrix} 2 + 3 + 0 \\ 2 + (-3) + 0 \\ 2 + 3 + 1 \end{bmatrix} \\ &= \begin{bmatrix} 5 \\ -1 \\ 6 \end{bmatrix} \end{aligned}$$

The vector $[5, -1, 6]^T$ is the same vector as $[2, 3, 1]^T$ but expressed in a different basis; the standard basis.

In order to understand better, let us consider a simpler example in \mathbb{R}^2 . With the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\} = \{[1, 0]^T, [0, 1]^T\}$ we have the vector $[1, 2]^T$. This vector can be expressed in terms of the standard basis like this:

$$\mathbf{x} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 = 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

The values $\alpha_1 = 1$ and $\alpha_2 = 2$ are called *coordinates* of \mathbf{x} with respect to $\{\mathbf{e}_1, \mathbf{e}_2\}$.

Now, let us introduce a new, ordered basis.

$$\mathcal{Q} = \left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = (\mathbf{q}_1, \mathbf{q}_2)$$

Let us say we have a coordinate $\mathbf{y} = [2, 3]^T$ with respect to the basis \mathcal{Q} . What is \mathbf{y} in the standard basis? The basis \mathcal{Q} gives us a recipe for the vector:

$$\mathbf{y} = \alpha_1 \mathbf{q}_1 + \alpha_2 \mathbf{q}_2 = 2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2+3 \\ -2+0 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$$

If we want to go the other way, from a point in the standard basis to a point with coordinates relative to \mathcal{Q} , we set up a matrix equation like this, using $[5, -2]^T$ as an example:

$$\alpha_1 \mathbf{q}_1 + \alpha_2 \mathbf{q}_2 = \begin{bmatrix} 5 \\ -2 \end{bmatrix} \implies [\mathbf{q}_1 \mathbf{q}_2] \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix} \implies \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$$

The matrix in the last step is made up of the basis columns, so the columns are linearly independent, and the matrix is invertible. The inverse matrix is:

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

We can use this to solve the matrix equation (by multiplying it to the left hand side of the equation). As the following calculations shows, the solution gives us back the coordinates $[2, 3]^T$ with respect to \mathcal{Q} , the opposite of what we found above:

$$\begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ -2 \end{bmatrix}$$

$$= \begin{bmatrix} 0(5) + (-1)(-2) \\ 1(5) + 1(-2) \end{bmatrix}$$

$$= \begin{bmatrix} 0 + 2 \\ 5 - 2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

As we have seen, the matrix defined by the columns of the basis \mathcal{Q} gives a mapping from coordinates under \mathcal{Q} to standard coordinates. The inverse matrix gives a mapping from the standard coordinates to \mathcal{Q} . To borrow some notation from another text book, we call this matrix the 'change-of-coordinate' matrix P and denote the relevant basis as a subscript.

$$[\mathbf{q}_1 \ \mathbf{q}_2] = P_{\mathcal{Q}} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} : \text{maps from a } \mathcal{Q} \text{ vector to a standard vector}$$

$$P_{\mathcal{Q}}^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} : \text{maps from a standard vector to a } \mathcal{Q} \text{ vector}$$

Now to look at translation between two vector spaces with non-standard bases. Continuing with the basis \mathcal{Q} , we introduce the basis \mathcal{R} :

$$\mathcal{R} = \left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right) = (\mathbf{r}_1, \mathbf{r}_2)$$

As we have seen above, the change-of-coordinate matrix from \mathcal{R} to standard basis is:

$$P_{\mathcal{R}} = [\mathbf{r}_1 \ \mathbf{r}_2] = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

The change-of-coordinate from standard basis to \mathcal{R} basis is the inverse matrix:

$$P_{\mathcal{R}}^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$$

In order to take a vector in \mathcal{R} and express it in \mathcal{Q} we first translate it from \mathcal{R} to standard coordinates, and then into \mathcal{Q} . (Borrowing some additional notation from another book where we explicitly state the basis of a vector):

$$[\mathbf{x}]_{\mathcal{Q}} = P_{\mathcal{Q}}^{-1} P_{\mathcal{R}} [\mathbf{x}]_{\mathcal{R}}$$

This means the translation matrix between \mathcal{R} and \mathcal{Q} is this:

$$P_{\mathcal{Q} \leftarrow \mathcal{R}} = P_{\mathcal{Q}}^{-1} P_{\mathcal{R}} = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

The opposite translation:

$$P_{\mathcal{R} \leftarrow \mathcal{Q}} = P_{\mathcal{Q} \leftarrow \mathcal{R}}^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$$

Or, when calculated explicitly with matrices.

$$P_{\mathcal{R} \leftarrow \mathcal{Q}} = P_{\mathcal{R}}^{-1} P_{\mathcal{Q}} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$$

Back to the exercise. We have the transformation matrix \mathbf{A}_Φ which translates from the standard basis to some unspecified basis, which we will call D . In our borrowed notation, this becomes P_D^{-1} .

$$\mathbf{A}_\Phi = P_D^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \implies P_D = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & -0.5 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

We are asked to find the transformation matrix $\tilde{\mathbf{A}}_\Phi$ wrt. to the provided basis:

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

As seen, the basis makes up the columns that map from B to the standard basis.

$$P_B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 1 & 0 \end{bmatrix} \implies P_B^{-1} = \begin{bmatrix} 0 & -1 & 2 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

They want us to find the matrix that maps from D to B . Going via the standard basis, we need a mapping from standard basis and from the standard basis to B .

$$[\mathbf{x}]_D \xrightarrow{P_D} \mathbf{x} \xrightarrow{P_B^{-1}} [\mathbf{x}]_B$$

Note the correct order of operation:

$$[\mathbf{x}]_B = P_B^{-1} P_D [\mathbf{x}]_D$$

According to my reasoning, the resulting change of basis matrix should be:

$$\begin{aligned} \tilde{\mathbf{A}}_\Phi = P_B^{-1} P_D &= \begin{bmatrix} 0 & -1 & 2 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & -0.5 & 0 \\ -1 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0(0.5) + (-1)(0.5) + 2(-1) & 0(0.5) + (-1)(-0.5) + 2(0) & 0(0) + (-1)(0) + 2(1) \\ 0(0.5) + 1(0.5) + (-1)(-1) & 0(0.5) + 1(-0.5) + (-1)(0) & 0(0) + 1(0) + (-1)(1) \\ 1(0.5) + 0(0.5) + (-1)(-1) & 1(0.5) + 0(-0.5) + (-1)(0) & 1(0) + 0(0) + (-1)(1) \end{bmatrix} \\ &= \begin{bmatrix} 0 - 0.5 - 2 & 0 + 0.5 + 0 & 0 - 0 + 2 \\ 0 + 0.5 + 1 & 0 - 0.5 - 0 & 0 + 0 - 1 \\ 0.5 + 0 + 1 & 0.5 - 0 - 0 & 0 + 0 - 1 \end{bmatrix} \\ &= \begin{bmatrix} -2.5 & 0.5 & 2 \\ 1.5 & -0.5 & -1 \\ 1.5 & 0.5 & -1 \end{bmatrix} \end{aligned}$$

2.20. We are given two ordered bases in \mathbb{R}^2 .

$$B = (\mathbf{b}_1, \mathbf{b}_2) = \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right), \quad B' = (\mathbf{b}'_1, \mathbf{b}'_2) = \left(\begin{bmatrix} 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)$$

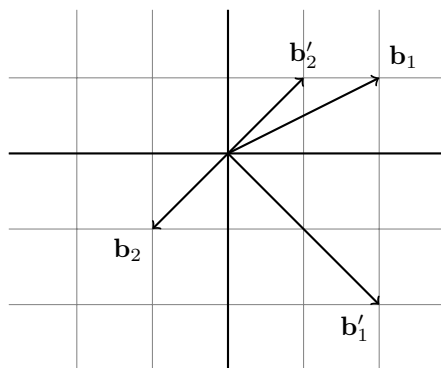
a.

Showing that both matrices have linearly independent columns, which means they are bases in \mathbb{R}^2 .

$$[\mathbf{b}_1 \ \mathbf{b}_2] = \begin{bmatrix} 2 & -1 \\ 1 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$[\mathbf{b}'_1 \ \mathbf{b}'_2] = \begin{bmatrix} 2 & 1 \\ -2 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Drawing the basis vectors.



b.

As in the earlier example, $P_{B'} = [\mathbf{b}'_1 \ \mathbf{b}'_2]$ maps a coordinate vector in $[\mathbf{x}]_{B'}$ back to the standard basis. The inverse of $P_B = [\mathbf{b}_1 \ \mathbf{b}_2]$ will map from the standard basis to $[\mathbf{x}]_B$. Finding the inverse:

$$P_B^{-1} = \begin{bmatrix} 2 & -1 \\ 1 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ 1 & -2 \end{bmatrix}$$

In order to map a vector from B' to B we need the following operations: $[\mathbf{x}]_B = P_B^{-1}P_{B'}[\mathbf{x}]_{B'}$. We can calculate the matrix.

$$\begin{aligned} \mathbf{P}_1 &= P_B^{-1}P_{B'} = \begin{bmatrix} 1 & -1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1(2) + (-1)(-2) & 1(1) + (-1)(1) \\ 1(2) + (-2)(-2) & 1(1) + (-2)(1) \end{bmatrix} \\ &= \begin{bmatrix} 2+2 & 1-1 \\ 2+4 & 1-2 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 0 \\ 6 & -1 \end{bmatrix} \end{aligned}$$

c.

We have the following ordered basis in \mathbb{R}^3 .

$$C = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3) = \left(\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right)$$

(i) Verifying that C is a basis. This is done by checking that the vectors are linearly independent. This can be verified by checking that the RREF is the identity matrix.

$$[\mathbf{c}_1 \ \mathbf{c}_2 \ \mathbf{c}_3] = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ -1 & 2 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(ii) The standard basis of \mathbb{R}^3 is $C' = (\mathbf{c}'_1, \mathbf{c}'_2, \mathbf{c}'_3)$. We are going to find the matrix \mathbf{P}_2 that performs a basis change from C to C' .

As seen earlier, a matrix where the columns make up a basis also gives us a recipe for mapping back to the standard basis.

$$\mathbf{P}_2 = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ -1 & 2 & -1 \end{bmatrix}$$

d.

We have a homomorphism (linear mapping) $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$, such that

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

Not stated, but I am assuming that these are the same bases as used earlier. Since Φ is a linear mapping, we can write:

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

Adding both these together, we get:

$$\begin{aligned} 2\Phi(\mathbf{b}_1) &= 2\mathbf{c}_1 + 4\mathbf{c}_3 \\ \Phi(\mathbf{b}_1) &= \mathbf{c}_1 + 2\mathbf{c}_3 \end{aligned}$$

Now we can find $\Phi(\mathbf{b}_2)$.

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

So, to sum up:

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

From this, we can read what the transformation matrix is (as shown in **Example 2.21** in the text book).

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

e.

Next we will find the transformation matrix \mathbf{A}' of Φ wrt. the bases B' and C' . When \mathbf{T} is the matrix that maps from C' to C and \mathbf{S} is the matrix that maps from B' to B , then

$$\mathbf{A}' = \mathbf{T}^{-1}\mathbf{A}_\Phi\mathbf{S}$$

as per **Theorem 2.20**. Starting with \mathbf{T} , which is NOT the matrix \mathbf{P}_2 we found earlier, which maps from C to C' ! Therefore, in our case, $\mathbf{T}^{-1} = \mathbf{P}_2$. so we need to find the inverse of that matrix.

The matrix \mathbf{S} is just \mathbf{P}_1 , so we can set up the entire calculation. (Not showing the matrix multiplications).

$$\begin{aligned} \mathbf{A}' &= \mathbf{T}^{-1}\mathbf{A}_\Phi\mathbf{S} \\ &= \mathbf{P}_2\mathbf{A}_\Phi\mathbf{P}_1 \\ &= \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 6 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 6 & -1 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 2 \\ -10 & 3 \\ 12 & -4 \end{bmatrix} \end{aligned}$$

f.

We have the following vector:

$$\mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}_{B'}$$

(i) Finding the corresponding vector expressed in B basis. This is just using matrix \mathbf{P}_1 :

$$\mathbf{P}_1[\mathbf{x}]_{B'} = \begin{bmatrix} 4 & 0 \\ 6 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix}_{B'} = \begin{bmatrix} 8 \\ 9 \end{bmatrix}_B$$

(ii) Now we translate this point into coordinates given by C using the transformation matrix \mathbf{A}_Φ .

$$\mathbf{A}_\Phi[\mathbf{x}]_B = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 8 \\ 9 \end{bmatrix}_B = \begin{bmatrix} -1 \\ 9 \\ 7 \end{bmatrix}_C$$

(iii) We translate the point into C' coordinates, which is done with \mathbf{P}_2 .

$$\mathbf{P}_2[\mathbf{x}]_C = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 9 \\ 7 \end{bmatrix}_C = \begin{bmatrix} 6 \\ -11 \\ 12 \end{bmatrix}_{C'}$$

(iv) Finally we do all the calculations in one step using the translation matrix \mathbf{A}' .

$$\mathbf{A}'[\mathbf{x}]_{B'} = \begin{bmatrix} 0 & 2 \\ -10 & 3 \\ 12 & -4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix}_{B'} = \begin{bmatrix} 6 \\ -11 \\ 12 \end{bmatrix}_{C'}$$

Chapter 3: Analytic Geometry - Exercises

3.1. We are going to verify that

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 - (x_1y_2 + x_2y_1) + 2x_2y_2$$

is an inner product, i.e. it is *symmetric*, *linear* and *positive definite*.

(Symmetry)

$$\begin{aligned}\langle \mathbf{x}, \mathbf{y} \rangle &= x_1y_1 - (x_1y_2 + x_2y_1) + 2x_2y_2 \\ &= y_1x_1 - (y_1x_2 + y_2x_1) + 2y_2x_2 \\ &= \langle \mathbf{y}, \mathbf{x} \rangle\end{aligned}$$

(Linearity)

$$\begin{aligned}\langle \alpha\mathbf{x} + \beta\mathbf{z}, \mathbf{y} \rangle &= (\alpha x_1 + \beta z_1)y_1 - ((\alpha x_1 + \beta z_1)y_2 + (\alpha x_2 + \beta z_2)y_1) + 2((\alpha x_2 + \beta z_2)y_2) \\ &= \alpha x_1y_1 + \beta z_1y_1 - (\alpha x_1y_2 + \beta z_1y_2 + \alpha x_2y_1 + \beta z_2y_1) + 2\alpha x_2y_2 + 2\beta z_2y_2 \\ &= \alpha x_1y_1 - (\alpha x_1y_2 + \alpha x_2y_1) + 2\alpha x_2y_2 + \beta z_1y_1 - (\beta z_1y_2 + \beta z_2y_1) + 2\beta z_2y_2 \\ &= \alpha(x_1y_1 - (x_1y_2 + x_2y_1) + 2x_2y_2) + \beta(z_1y_1 - (z_1y_2 + z_2y_1) + 2z_2y_2) \\ &= \alpha\langle \mathbf{x}, \mathbf{y} \rangle + \beta\langle \mathbf{z}, \mathbf{y} \rangle\end{aligned}$$

By the symmetric property, the opposite linear argument holds.

(Positive Definite)

Assume $\mathbf{x} \neq \mathbf{0}$, that is either $x_1 \neq 0$ or $x_2 \neq 0$, then

$$\begin{aligned}\langle \mathbf{x}, \mathbf{x} \rangle &= x_1^2 - (x_1x_2 + x_2x_1) + 2x_2^2 \\ &= x_1^2 - 2x_1x_2 + 2x_2^2 \\ &= x_1^2 - 2x_1x_2 + x_2^2 + x_2^2 \\ &= (x_1 - x_2)^2 + x_2^2 > 0\end{aligned}$$

As the sum of two positive values, this is always a positive value and strictly larger than 0 because we assumed that $\mathbf{x} \neq \mathbf{0}$. By setting in the null vector, we get:

$$\langle \mathbf{0}, \mathbf{0} \rangle = (0 - 0)^2 + 0^2 = 0$$

This shows that the operation is positive definite. All properties are satisfied, so $\langle \cdot, \cdot \rangle$ is an inner product.

3.2. We will check if the following operation is an inner product.

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} \mathbf{y} = \mathbf{x}^T \begin{bmatrix} 2y_1 \\ y_1 + 2y_2 \end{bmatrix} = 2x_1y_1 + y_1x_2 + 2y_2x_2$$

This is NOT an inner product. It breaks symmetry:

$$\langle \mathbf{y}, \mathbf{x} \rangle = 2x_1y_1 + x_1y_2 + 2y_2x_2$$

(since $y_1x_2 \neq x_1y_2$). And it does not have the positive definite property, which can be shown with the counterexample $\mathbf{x}_c = [-1, -1]^T$ and $\mathbf{y}_c = [1, 1]^T$. Then:

$$\langle \mathbf{x}_c, \mathbf{y}_c \rangle = 2(-1)(1) + (1)(-1) + 2(1)(-1) = -2 - 1 - 2 = -5$$

3.3. Computing the distance between two vectors, where distance is $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle}$.

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

a.

Using the usual dot product. Start by calculating the difference vector.

$$\mathbf{x} - \mathbf{y} = \begin{bmatrix} 1 - (-1) \\ 2 - (-1) \\ 3 - 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$$

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle} = \sqrt{2^2 + 3^2 + 3^2} = \sqrt{4 + 9 + 9} = \sqrt{22}$$

b.

Calculating with $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{A} \mathbf{y}$.

$$\begin{aligned} \|\mathbf{x} - \mathbf{y}\| &= \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle} = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{A} (\mathbf{x} - \mathbf{y})} \\ &= \left([2 \ 3 \ 3] \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} \right)^{\frac{1}{2}} \\ &= \left([2 \ 3 \ 3] \begin{bmatrix} 7 \\ 8 \\ 3 \end{bmatrix} \right)^{\frac{1}{2}} \\ &= \sqrt{14 + 24 + 9} = \sqrt{47} \end{aligned}$$

3.4. Finding the angle. We have the vectors:

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

General formula for calculating the angle between two vectors \mathbf{x} and \mathbf{y} :

$$\cos \omega = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad (\spadesuit)$$

a.

Calculating the angle with the standard dot product. First we find the lengths of the vectors.

$$\|\mathbf{x}\| = \sqrt{1^2 + 2^2} = \sqrt{5}$$

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

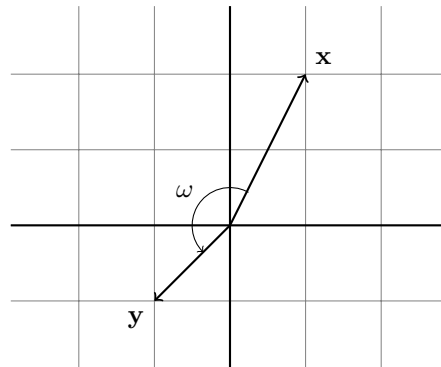
The dot product.

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

Applying to the formula ().

$$\cos \omega = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} = \frac{-3}{\sqrt{5}\sqrt{2}} \implies \omega = 2.819842$$

The angle corresponds to around 161.5° . This can be seen in the following drawing, where the angle is close to 180° .



b.

Calculating the angle when

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{B} \mathbf{y}.$$

Finding the lengths.

$$\|\mathbf{x}\|^2 = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix} = 18 \implies \|\mathbf{x}\| = \sqrt{18} = 3\sqrt{2}$$

$$\|\mathbf{y}\|^2 = \begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} -3 \\ -4 \end{bmatrix} = 7 \implies \|\mathbf{y}\| = \sqrt{7}$$

Finding the inner product.

$$\langle \mathbf{x}, \mathbf{y} \rangle = [1 \ 2] \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \end{bmatrix} = [1 \ 2] \begin{bmatrix} -3 \\ -4 \end{bmatrix} = (1)(-3) + (2)(-4) = -3 - 8 = -11$$

Applying the formula (♠).

$$\cos \omega = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} = \frac{-11}{3\sqrt{2}\sqrt{7}} \implies \omega = 2.94104$$

With this inner product, the angle is around 168.5° .

3.5. We have a subspace $U \subset \mathbb{R}^5$ and $\mathbf{x} \in \mathbb{R}^5$ given by:

$$U = \text{span} \left(\begin{bmatrix} 0 \\ -1 \\ 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -3 \\ 5 \\ 0 \\ 7 \end{bmatrix} \right), \quad \mathbf{x} = \begin{bmatrix} -1 \\ -9 \\ -1 \\ 4 \\ 1 \end{bmatrix}$$

a.

Finding the orthogonal projection π_U from \mathbf{x} to U . First an important observation. The columns that span U are not linearly independent.

$$\begin{bmatrix} 0 & 1 & -3 & -1 \\ -1 & -3 & 4 & -3 \\ 2 & 1 & 1 & 5 \\ 0 & -1 & 2 & 0 \\ 2 & 2 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The fourth column is a linear combination of the first three columns. This means that U spans a 3-dimensional space, and the proper basis will only be the first 3 columns.

Following the notation in the text book, let us call the vectors that span $U = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ and the matrix with these vectors as its column B . From equation (3.58) in the text book, we get the following formula for the projection (skipping the intermediate matrix multiplications):

$$\pi_U(\mathbf{x}) = B(B^T B)^{-1} B^T \mathbf{x}$$

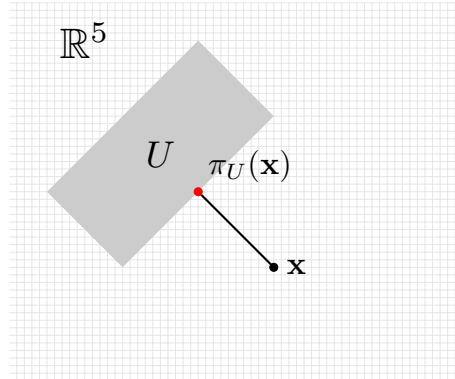
$$= \begin{bmatrix} 0.4762 & -0.2857 & 0.1905 & -0.1429 & -0.3333 \\ -0.2857 & 0.6825 & 0.0635 & 0.2857 & -0.2222 \\ 0.1905 & 0.0635 & 0.9206 & 0.1429 & 0.1111 \\ -0.1429 & 0.2857 & 0.1429 & 0.1429 & 0 \\ -0.3333 & -0.2222 & 0.1111 & 0 & 0.7778 \end{bmatrix} \begin{bmatrix} -1 \\ -9 \\ -1 \\ 4 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ -5 \\ -1 \\ -2 \\ 3 \end{bmatrix}$$

Not showing the calculations, but we can verify that $\pi_U(\mathbf{x}) \in U$ by finding the RREF of $[\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{b}_3 \ \pi_U(\mathbf{x})]$. It turns out that:

$$\pi_U(\mathbf{x}) = -3\mathbf{b}_1 + 4\mathbf{b}_2 + \mathbf{b}_3.$$

For a set $U \subset \mathbb{R}^5$ and some point $\mathbf{x} \notin U$ we have found the projection $\pi_U(\mathbf{x})$ marked in red, which is the point in U closest to the point \mathbf{x} . Here is an illustration:



b.

Calculating the distance $d(\mathbf{x}, U)$, which amounts to calculating the shortest distance $d(\mathbf{x}, \pi_U(\mathbf{x}))$. Calculating the difference vector:

$$\mathbf{x} - \pi_U(\mathbf{x}) = \begin{bmatrix} -1 \\ -9 \\ -1 \\ 4 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ -5 \\ -1 \\ -2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \\ 0 \\ 6 \\ -2 \end{bmatrix}$$

$$\begin{aligned} d(\mathbf{x}, \pi_U(\mathbf{x})) &= \|\mathbf{x} - \pi_U(\mathbf{x})\| = \sqrt{\langle \mathbf{x} - \pi_U(\mathbf{x}), \mathbf{x} - \pi_U(\mathbf{x}) \rangle} \\ &= \sqrt{(-2)^2 + (-4)^2 + 0^2 + 6^2 + (-2)^2} \\ &= \sqrt{4 + 16 + 36 + 4} \\ &= \sqrt{60} \\ &= 2\sqrt{15} \end{aligned}$$

3.6. Working in \mathbb{R}^3 with the standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{A} \mathbf{y} = \mathbf{x}^T \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \mathbf{y}.$$

a.

We are going to determine the orthogonal projection of $\pi_U(\mathbf{e}_2)$ onto the plane in \mathbb{R}^3 defined as

$$U = \text{span}(\mathbf{e}_1, \mathbf{e}_3) = \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right)$$

Following the general guidelines from the text book. The projection we need must satisfy the following:

$$\begin{aligned} \pi_U(\mathbf{e}_2) &= \lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_3 \\ &= [\mathbf{e}_1 \quad \mathbf{e}_3] \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \\ &= \mathbf{B} \lambda, \end{aligned}$$

such that the λ_i ensures that $\pi_U(\mathbf{e}_2)$ is closest to \mathbf{e}_2 . In other words, the vector $\mathbf{e}_2 - \pi_U(\mathbf{e}_2)$ must be orthogonal to all basis vectors of U . This gives the following conditions:

$$\begin{aligned} \langle \mathbf{e}_1, \mathbf{e}_2 - \pi_U(\mathbf{e}_2) \rangle &= 0 \\ \langle \mathbf{e}_3, \mathbf{e}_2 - \pi_U(\mathbf{e}_2) \rangle &= 0 \end{aligned}$$

By linearity of the inner product.

$$\begin{aligned} \langle \mathbf{e}_1, \mathbf{e}_2 \rangle - \langle \mathbf{e}_1, \pi_U(\mathbf{e}_2) \rangle &= 0 \\ \langle \mathbf{e}_3, \mathbf{e}_2 \rangle - \langle \mathbf{e}_3, \pi_U(\mathbf{e}_2) \rangle &= 0 \end{aligned}$$

Writing out the matrix equations.

$$\begin{aligned} \mathbf{e}_1^T \mathbf{A} \mathbf{e}_2 - \mathbf{e}_1^T \mathbf{A} \pi_U(\mathbf{e}_2) &= 0 \\ \mathbf{e}_3^T \mathbf{A} \mathbf{e}_2 - \mathbf{e}_3^T \mathbf{A} \pi_U(\mathbf{e}_2) &= 0 \end{aligned}$$

$$\begin{aligned} \mathbf{e}_1^T \mathbf{A} \mathbf{e}_2 &= \mathbf{e}_1^T \mathbf{A} \pi_U(\mathbf{e}_2) \\ \mathbf{e}_3^T \mathbf{A} \mathbf{e}_2 &= \mathbf{e}_3^T \mathbf{A} \pi_U(\mathbf{e}_2) \end{aligned}$$

We can simplify a bit by reducing some of the matrices and vectors into numbers.

$$\begin{aligned}\mathbf{e}_1^T \mathbf{A} \mathbf{e}_2 &= [1 \ 0 \ 0] \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\ &= [1 \ 0 \ 0] \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \\ &= 1\end{aligned}$$

$$\begin{aligned}\mathbf{e}_3^T \mathbf{A} \mathbf{e}_2 &= [0 \ 0 \ 1] \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\ &= [0 \ 0 \ 1] \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \\ &= -1\end{aligned}$$

The projection point is unknown, so we will call the coordinates x_1, x_2 and x_3 .

$$\begin{aligned}\mathbf{e}_1^T \mathbf{A} \pi_U(\mathbf{e}_2) &= [1 \ 0 \ 0] \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= [1 \ 0 \ 0] \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 - x_3 \\ -x_2 + 2x_3 \end{bmatrix} \\ &= 2x_1 + x_2\end{aligned}$$

$$\begin{aligned}\mathbf{e}_3^T \mathbf{A} \pi_U(\mathbf{e}_2) &= [0 \ 0 \ 1] \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= [0 \ 0 \ 1] \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 - x_3 \\ -x_2 + 2x_3 \end{bmatrix} \\ &= -x_2 + 2x_3\end{aligned}$$

Recalling the equations.

$$\begin{aligned}\mathbf{e}_1^T \mathbf{A} \mathbf{e}_2 &= \mathbf{e}_1^T \mathbf{A} \pi_U(\mathbf{e}_2) \\ \mathbf{e}_3^T \mathbf{A} \mathbf{e}_2 &= \mathbf{e}_3^T \mathbf{A} \pi_U(\mathbf{e}_2)\end{aligned}$$

Replacing the known quantities.

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

We can add a third condition, since the plane U is spanned by \mathbf{e}_1 and \mathbf{e}_3 , it will always be 0 in \mathbf{e}_2 . The same must be true for the projected point on the plane, which means that $x_2 = 0$. This gives us.

$$\begin{aligned}1 &= 2x_1 \\ -1 &= 2x_3\end{aligned}$$

When solved for x_1 and x_3 .

$$\begin{aligned}x_1 &= \frac{1}{2} \\ x_3 &= -\frac{1}{2}\end{aligned}$$

Hence we know that:

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

b.

Calculating the distance $d(\mathbf{e}_2, \pi_U(\mathbf{e}_2))$. Finding the difference vector.

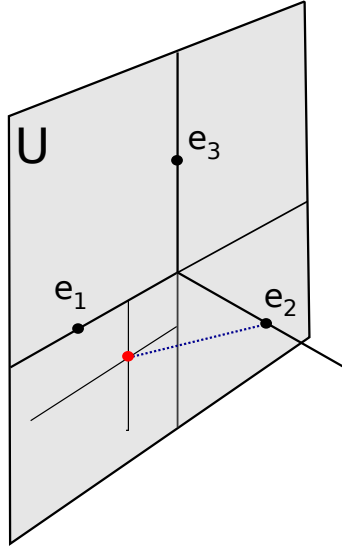
$$\mathbf{e}_2 - \pi_U(\mathbf{e}_2) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 1 \\ \frac{1}{2} \end{bmatrix}$$

Calculating the distance with the inner product.

$$\begin{aligned}d(\mathbf{e}_2, \pi_U(\mathbf{e}_2)) &= \|\mathbf{e}_2 - \pi_U(\mathbf{e}_2)\| = \sqrt{\langle \mathbf{e}_2 - \pi_U(\mathbf{e}_2), \mathbf{e}_2 - \pi_U(\mathbf{e}_2) \rangle} \\ &= \left(\begin{bmatrix} -\frac{1}{2} & 1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} \\ 1 \\ \frac{1}{2} \end{bmatrix} \right)^{\frac{1}{2}} \\ &= \left(\begin{bmatrix} -\frac{1}{2} & 1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right)^{\frac{1}{2}} = \sqrt{1} = 1\end{aligned}$$

c.

Illustration of the situation. U is a plane going along the \mathbf{e}_1 and \mathbf{e}_3 bases. The projected point on $\pi_U(\mathbf{e}_2) \in U$ is marked as a red point. The 'shortest' distance from \mathbf{e}_2 to the plane is distorted because we are using a special kind of inner product.



3.7. Recalling a term. *Endomorphism:* $\pi : V \rightarrow V$ is linear. Let V be a vector space and π and endomorphism of V , and id_V is the identity endomorphism on V .

a.

Prove that π is a projection if and only if $id_V - \pi$ is a projection.

Proof.

\Rightarrow) Assuming π is a projection. That means that for any $\mathbf{x} \in V$, then $\pi(\pi(\mathbf{x})) = \pi(\mathbf{x})$. To show that $id_V(\mathbf{x}) - \pi(\mathbf{x})$ is a projection, we have to show that

$$id_V(id_V(\mathbf{x}) - \pi(\mathbf{x})) - \pi(id_V(\mathbf{x}) - \pi(\mathbf{x})) = id_V(\mathbf{x}) - \pi(\mathbf{x}).$$

Doing the calculations:

$$\begin{aligned} & id_V(id_V(\mathbf{x}) - \pi(\mathbf{x})) - \pi(id_V(\mathbf{x}) - \pi(\mathbf{x})) \\ \stackrel{\text{Linearity}}{=} & id_V(id_V(\mathbf{x})) - id_V(\pi(\mathbf{x})) - [\pi(id_V(\mathbf{x})) - \pi(\pi(\mathbf{x}))] \\ \stackrel{\text{Identity}}{=} & id_V(\mathbf{x}) - \pi(\mathbf{x}) - [\pi(\mathbf{x}) - \pi(\pi(\mathbf{x}))] \\ \stackrel{\text{Assumpt.}}{=} & id_V(\mathbf{x}) - \pi(\mathbf{x}) - \pi(\mathbf{x}) + \pi(\mathbf{x}) \\ = & id_V(\mathbf{x}) - \pi(\mathbf{x}) \end{aligned}$$

This shows that $id_V - \pi$ is a projection when π is a projection.

\Leftrightarrow) Assuming $id_V(\mathbf{x}) - \pi(\mathbf{x})$ is a projection. This means that for any $\mathbf{x} \in V$:

$$id_V(id_V(\mathbf{x}) - \pi(\mathbf{x})) - \pi(id_V(\mathbf{x}) - \pi(\mathbf{x})) = id_V(\mathbf{x}) - \pi(\mathbf{x}).$$

By subtracting the right hand side, we get:

$$id_V(id_V(\mathbf{x}) - \pi(\mathbf{x})) - \pi(id_V(\mathbf{x}) - \pi(\mathbf{x})) - id_V(\mathbf{x}) + \pi(\mathbf{x}) = \mathbf{0}$$

By the linearity property and the identity property as shown above, we get:

$$id_V(\mathbf{x}) - \pi(\mathbf{x}) - [\pi(\mathbf{x}) - \pi(\pi(\mathbf{x}))] - id_V(\mathbf{x}) + \pi(\mathbf{x}) = \mathbf{0}$$

$$\cancel{id_V(\mathbf{x}) - \pi(\mathbf{x})} - \pi(\mathbf{x}) + \pi(\pi(\mathbf{x})) - \cancel{id_V(\mathbf{x}) + \pi(\mathbf{x})} = \mathbf{0}$$

Which leaves us with:

$$\pi(\pi(\mathbf{x})) = \pi(\mathbf{x})$$

This shows that π is a projection when $id_v - \pi$ is a projection. By implications both ways, we have proved that the statements are equivalent. ■

b.

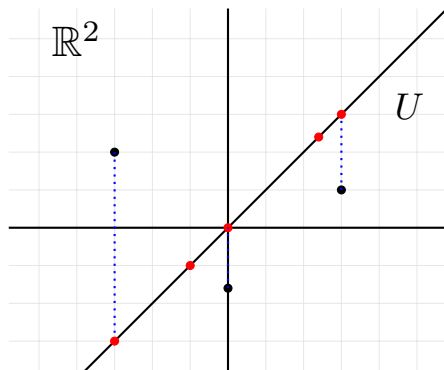
Assuming π is a projection. Calculating $\text{Im}(id_V - \pi)$ and $\text{ker}(id_V - \pi)$ as a function of $\text{Im}(\pi)$ and $\text{ker}(\pi)$.

Let's assume that U is some subspace of V and that π is a projection onto that subspace. Let us also consider a simplified example to make this more concrete. In \mathbb{R}^2 we have the subspace U given by $y = x$ as shown in this image.

For any point $\mathbf{x} \in U$ the projection will simply return \mathbf{x} . For any point $\mathbf{x} \notin U$ the projection will be a point on the 'boundary' of U . This means that $\text{Im}(\pi) = U$.

Any point on the y -axis will be projected down to the origin, so we can express the kernel of π as:

$$\text{ker}(\pi) = \left\{ \begin{bmatrix} 0 \\ y \end{bmatrix} : y \in \mathbb{R} \right\}$$



Now if we have some point $x \in U$ such as $(1, 1)$, then id_V will return $(1, 1)$ and so will π . In fact, $id_V - \pi$ will map any point in U to $\mathbf{0}$. For any point outside U such as $(2, 5)$, the projection will map it down to $(2, 2)$. The difference $id_V - \pi$ will be $(0, 3)$ which is in the kernel of π .

A point in the kernel of π , such as $(0, 2)$ will be projected to the origin, and the difference $id_V - \pi$ will be $(0, 2)$, which is still in the kernel of π .

With this simplified example in mind, we can move on to the more general terms. Now U is some general subspace of V .

Image $\text{Im}(id_V - \pi) = \ker(\pi)$

Proof.

⊂) Assume $\mathbf{x} \in \text{Im}(id_V - \pi)$. There are two cases we must check.

(i) Assume further that $\mathbf{x} \in U$, then $\pi(\mathbf{x}) = \mathbf{x}$. In this case

$$id_V(\mathbf{x}) - \pi(\mathbf{x}) = \mathbf{x} - \mathbf{x} = \mathbf{0}$$

which shows that $\mathbf{x} \in \ker(\pi)$ since $\mathbf{0} \in U$ and $\pi(\mathbf{0}) = \mathbf{0}$. This means $\text{Im}(id_V - \pi) \subset \ker(\pi)$ when $\mathbf{x} \in U$.

(ii) Now assume $\mathbf{x} \notin U$.

$$id_V(\mathbf{x}) - \pi(\mathbf{x}) = \mathbf{x} - \pi(\mathbf{x})$$

If we take the projection of this point:

$$\pi(id_V(\mathbf{x}) - \pi(\mathbf{x})) = \pi(\mathbf{x}) - \pi(\pi(\mathbf{x})) = \pi(\mathbf{x}) - \pi(\mathbf{x}) = \mathbf{0}$$

which shows that $\mathbf{x} \in \ker(\pi)$ so $\text{Im}(id_V - \pi) \subset \ker(\pi)$ when $\mathbf{x} \notin U$.

Both of these together show that $\text{Im}(id_V - \pi) \subset \ker(\pi)$ for all $x \in V$.

⊃) Assume now that $\mathbf{x} \in \ker(\pi)$, so $\pi(\mathbf{x}) = \mathbf{0}$.

$$\mathbf{x} = id_V(\mathbf{x}) - \mathbf{0} = id_V(\mathbf{x}) - \pi(\mathbf{x})$$

which shows that $\mathbf{x} \in \text{Im}(id_V - \pi)$ and therefore $\ker(\pi) \subset \text{Im}(id_V - \pi)$. By inclusion both ways, we can conclude that $\text{Im}(id_V - \pi) = \ker(\pi)$. ■

Kernel $\ker(id_V - \pi) = \text{Im}(\pi)$

Proof.

⊂) Assume $\mathbf{x} \in \ker(id_V - \pi)$, so $id_V(\mathbf{x}) - \pi(\mathbf{x}) = \mathbf{0}$. Then the following must be true:

$$id_V(\mathbf{x}) - \pi(\mathbf{x}) = \mathbf{0} \implies id_V(\mathbf{x}) = \pi(\mathbf{x}) \implies \mathbf{x} = \pi(\mathbf{x})$$

and the projection $\pi(\mathbf{x}) = \mathbf{x}$ only when $\mathbf{x} \in \text{Im}(\pi)$ which shows that $\ker(id_V - \pi) \subset \text{Im}(\pi)$.

⊃) Assume now that $\mathbf{x} \in \text{Im}(\pi)$ which means $\pi(\mathbf{x}) = \mathbf{x}$. The reverse argument shows:

$$\mathbf{x} = \pi(\mathbf{x}) \implies id_V(\mathbf{x}) = \pi(\mathbf{x}) \implies id_V(\mathbf{x}) - \pi(\mathbf{x}) = \mathbf{0}$$

so $\mathbf{x} \in \ker(id_V - \pi)$ which shows $\text{Im}(\pi) \subset \ker(id_V - \pi)$.

By inclusion both ways we have proved that $\ker(id_V - \pi) = \text{Im}(\pi)$. ■

3.8. Using Gram-Schmidt to turn the basis $B = (\mathbf{b}_1, \mathbf{b}_2)$ into an ONB (orthonormal basis) C .

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

We start by making the columns orthogonal, which they aren't since $\mathbf{b}_1^T \mathbf{b}_2 = 1$. Set $\mathbf{v}_1 = \mathbf{b}_1$. Then we can construct the orthogonal vector from \mathbf{b}_2 as follows:

$$\mathbf{v}_2 = \mathbf{b}_2 - \frac{\mathbf{b}_1^T \mathbf{b}_2}{\mathbf{b}_1^T \mathbf{b}_1} \mathbf{b}_1$$

Some intermediary calculations:

$$\mathbf{b}_1^T \mathbf{b}_2 = (1)(-1) + (1)(2) + 0 = -1 + 2 = 1$$

$$\mathbf{b}_1^T \mathbf{b}_1 = (1)(1) + (1)(1) + (1)(1) = 1 + 1 + 1 = 3$$

We get:

$$\mathbf{v}_2 = \mathbf{b}_2 - \frac{\mathbf{b}_1^T \mathbf{b}_2}{\mathbf{b}_1^T \mathbf{b}_1} \mathbf{b}_1$$

$$= \mathbf{b}_2 - \frac{1}{3} \mathbf{b}_1$$

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

$$= \begin{bmatrix} -4/3 \\ 5/3 \\ -1/3 \end{bmatrix}$$

The vectors \mathbf{v}_1 and \mathbf{v}_2 are orthogonal, which we can see by calculating the inner product.

$$\mathbf{v}_1^T \mathbf{v}_2 = (1)(-4/3) + (1)(5/3) + (1)(-1/3) = 0$$

The final step is to make the vectors orthonormal. This is done by scaling them down.

$$\mathbf{c}_1 = \frac{1}{\|\mathbf{v}_1\|} \mathbf{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{c}_2 = \frac{1}{\|\mathbf{v}_2\|} \mathbf{v}_2 = \frac{1}{\sqrt{42}} \begin{bmatrix} -4 \\ 5 \\ -1 \end{bmatrix}$$

Where:

$$\|\mathbf{v}_2\| = \sqrt{(-4/3)^2 + (5/3)^2 + (-1/3)^2} = \sqrt{42/9} = \sqrt{42}/3 \implies \frac{1}{\|\mathbf{v}_2\|} = \frac{3}{\sqrt{42}}$$

3.9. Let $n \in \mathbb{N}$ and $x_1, \dots, x_n > 0$ such that $x_1 + \dots + x_n = 1$. We will need the Cauchy-Schwarz inequality (3.17):

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|.$$

a.

Showing that

$$n \leq \sum_{i=1}^n x_i^2.$$

We choose the following vectors:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

The inner product becomes:

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle &= (x_1)(1) + (x_2)(1) + \dots + (x_n)(1) \\ &= x_1 + x_2 + \dots + x_n \\ &= 1 \end{aligned}$$

where the last equality holds due to the property of the x_i values. The norms become:

$$\begin{aligned} \|\mathbf{x}\| &= \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \\ \|\mathbf{y}\| &= \sqrt{1^2 + 1^2 + \dots + 1^2} = \sqrt{n} \end{aligned}$$

By the Cauchy-Schwarz inequality (don't need the absolute value since all values are positive).

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$$

From our calculations, this means:

$$1 \leq \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \sqrt{n}$$

Squaring both sides (inequality remains unchanged as both sides are positive).

$$\begin{aligned} 1 &\leq (x_1^2 + x_2^2 + \dots + x_n^2) n \\ 1 &\leq n \sum_{i=1}^n x_i^2 \end{aligned}$$

Dividing both sides with n .

$$\frac{1}{n} \leq \sum_{i=1}^n x_i^2$$

This is the result we wanted to show.

b.

Showing that

$$n^2 \leq \sum_{i=1}^n \frac{1}{x_i}.$$

We choose the following vectors:

$$\mathbf{x} = \begin{bmatrix} \sqrt{x_1} \\ \vdots \\ \sqrt{x_n} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \frac{1}{\sqrt{x_1}} \\ \vdots \\ \frac{1}{\sqrt{x_n}} \end{bmatrix}$$

Dot product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \left(\sqrt{x_1} \cdot \frac{1}{\sqrt{x_1}} \right) + \dots + \left(\sqrt{x_n} \cdot \frac{1}{\sqrt{x_n}} \right) = 1 + \dots + 1 = n$$

Norms.

$$\|\mathbf{x}\| = \sqrt{(\sqrt{x_1})^2 + \dots + (\sqrt{x_n})^2} = \sqrt{x_1 + \dots + x_n} = \sqrt{1} = 1$$

$$\|\mathbf{y}\| = \sqrt{\left(\frac{1}{\sqrt{x_1}}\right)^2 + \dots + \left(\frac{1}{\sqrt{x_n}}\right)^2} = \sqrt{\frac{1}{x_1} + \dots + \frac{1}{x_n}}$$

Applying the Cauchy-Schwarz inequality.

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$$

In this case, we have found that we have:

$$n \leq \sqrt{\frac{1}{x_1} + \dots + \frac{1}{x_n}}$$

Squaring both sides.

$$n^2 \leq \frac{1}{x_1} + \dots + \frac{1}{x_n}$$

Which is what we wanted to show.

$$n^2 \leq \sum_{i=1}^n \frac{1}{x_i}$$

3.10. Rotating the following vectors by 30° .

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

In degrees 30° is 1/12th of a full circle. In radians this is

$$\theta = \frac{2\pi}{12} = \frac{\pi}{6}.$$

With this, we can define a standard rotation matrix R :

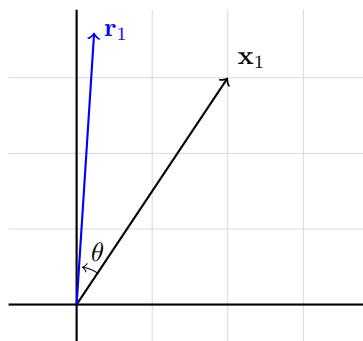
$$R = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$$

Finding the rotated vectors:

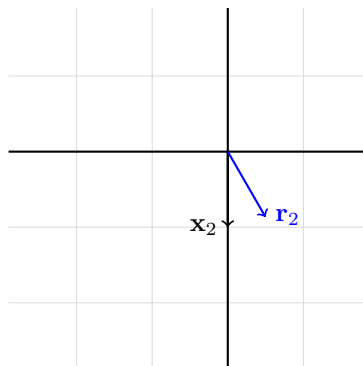
$$\mathbf{r}_1 = R\mathbf{x}_1 = \begin{bmatrix} 0.2320508 \\ 3.5980762 \end{bmatrix}$$

$$\mathbf{r}_2 = R\mathbf{x}_2 = \begin{bmatrix} 0.5 \\ -0.8660254 \end{bmatrix}$$

Plotting the rotation of \mathbf{x}_1 .



Plotting the rotation of \mathbf{x}_2 .



Chapter 4: Matrix Decompositions - Exercises

4.1. Calculating the determinant with Laplacian expansion across the first row of A .

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

$$\begin{aligned} \det(A) &= (1) \begin{vmatrix} 4 & 6 \\ 2 & 4 \end{vmatrix} - (3) \begin{vmatrix} 2 & 6 \\ 0 & 4 \end{vmatrix} + (5) \begin{vmatrix} 2 & 4 \\ 0 & 2 \end{vmatrix} \\ &= (1)(16 - 12) - (3)(8 - 0) + (5)(4 - 0) \\ &= 4 - 24 + 20 \\ &= 0 \end{aligned}$$

4.2. Finding an efficient way of computing the following determinant.

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

Using the property of determinants which say that adding a row to another does not change the determinant. We can do this and simplify the calculations.

$$A \stackrel{II \sim V}{\sim} \begin{bmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 2 \\ -2 & 0 & 2 & -1 & 2 \\ 2 & 0 & 0 & 1 & 1 \end{bmatrix} \stackrel{V \sim I}{\sim} \begin{bmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 2 \\ -2 & 0 & 2 & -1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{bmatrix} \stackrel{IV \sim +I}{\sim} \begin{bmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{bmatrix} \stackrel{III \sim -IV}{\sim} \begin{bmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{bmatrix}$$

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

$$= (2)(-1)(-1)((1)(1) - (-1)(2)) = (2)(-1)(-1)(3) = 6$$

Verified by calculating the original determinant on a computer.

4.3. Computing the eigenspace.

a.

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

Finding the characteristic polynomial.

$$\det(A - \lambda I_2) = \det\left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = \begin{vmatrix} 1-\lambda & 0 \\ 1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2$$

When setting this equal to 0, we see that the solution is $\lambda = 1$ with multiplicity 2. To find the eigenvectors, we solve the homogeneous system:

$$\begin{bmatrix} 1-\lambda & 0 \\ 1 & 1-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Setting in $\lambda = 1$.

$$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Writing this out as an augmented matrix gives the result:

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

So, x_2 is a free variable and $x_1 = 0$. The eigenvector is therefore:

$$\mathbf{x} = \begin{bmatrix} 0 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

for any $x_2 \in \mathbb{R}$. Let's set $x_2 = 1$ and check the answer. First the matrix calculation.

$$\begin{aligned} A\mathbf{x} &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1(0) + 0(1) \\ 1(0) + 1(1) \end{bmatrix} \\ &= \begin{bmatrix} 0 + 0 \\ 0 + 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

And, with the eigenvalue.

$$\lambda\mathbf{x} = (1) \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Hence, the eigenspace is spanned by:

$$\text{span}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right).$$

b.

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

Finding the characteristic polynomial.

$$\begin{aligned} \det(B - \lambda I_2) &= \det \left(\begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) = \begin{vmatrix} -2-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = (-2-\lambda)(1-\lambda) - 2 \cdot 2 \\ &= \lambda^2 + \lambda - 6 \implies (\lambda - 2)(\lambda + 3) \end{aligned}$$

The solutions are $\lambda_1 = 2$ and $\lambda_2 = -3$. Now we can find the eigenvectors by solving $(B - \lambda I)\mathbf{x} = \mathbf{0}$.

Using $\lambda_1 = 2$:

$$\begin{bmatrix} -4 & 2 & 0 \\ 2 & -1 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & -0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = 0.5x_2 \\ x_2 = x_2 \end{array} \implies \mathbf{x} = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$$

Verifying:

$$B\mathbf{x} = \begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \lambda\mathbf{x} = (2) \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Using $\lambda_2 = -3$

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = -2x_2 \\ x_2 = x_2 \end{array} \implies \mathbf{x} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

Verifying:

$$B\mathbf{x} = \begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ -3 \end{bmatrix}, \quad \lambda\mathbf{x} = (-3) \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ -3 \end{bmatrix}$$

From these results we have the following eigenspaces.

$$E_2 = \text{span} \left(\begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \right), \quad E_{-3} = \text{span} \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix} \right)$$

4.4. Computing eigenspaces for A .

$$A = \begin{bmatrix} 0 & -1 & 1 & 1 \\ -1 & 1 & -2 & 3 \\ 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{bmatrix} \implies (A - \lambda I_4) = \begin{bmatrix} -\lambda & -1 & 1 & 1 \\ -1 & 1 - \lambda & -2 & 3 \\ 2 & -1 & -\lambda & 0 \\ 1 & -1 & 1 & -\lambda \end{bmatrix}$$

Finding the characteristic polynomial. (Skipping the calculations).

$$\begin{aligned} \det(A - \lambda I_4) &= \begin{vmatrix} -\lambda & -1 & 1 & 1 \\ -1 & 1 - \lambda & -2 & 3 \\ 2 & -1 & -\lambda & 0 \\ 1 & -1 & 1 & -\lambda \end{vmatrix} = \lambda^4 - \lambda^3 - 3\lambda^2 + \lambda + 2 \\ &= (\lambda + 1)^2(\lambda - 1)(\lambda - 2) \end{aligned}$$

So the roots are: $\lambda_1 = -1$ (with multiplicity 2), $\lambda_2 = 1$ and $\lambda_3 = 2$. Finding the eigenvectors by solving $(A - \lambda I_4)\mathbf{x} = \mathbf{0}$.

Using $\lambda_1 = -1$ for $(A - \lambda I_4)$ in the augmented matrix $[(A - \lambda I_4) \mathbf{0}]$:

$$\begin{bmatrix} 1 & -1 & 1 & 1 & 0 \\ -1 & 2 & -2 & 3 & 0 \\ 2 & -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 1 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = 0 \\ x_2 = x_3 \\ x_3 = x_3 \\ x_4 = 0 \end{array} \implies \mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 0 & -1 & 1 & 1 \\ -1 & 1 & -2 & 3 \\ 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \end{bmatrix}, \quad \lambda\mathbf{x} = (-1) \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \end{bmatrix}$$

Even though the characteristic polynomial has algebraic multiplicity of 2, it only has one corresponding eigenvector, so the geometric multiplicity is 1.

Using $\lambda_2 = 1$ for $(A - \lambda I_4)$ in the augmented matrix $[(A - \lambda I_4) \mathbf{0}]$:

$$\begin{bmatrix} -1 & -1 & 1 & 1 & 0 \\ -1 & 0 & -2 & 3 & 0 \\ 2 & -1 & -1 & 0 & 0 \\ 1 & -1 & 1 & -1 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = x_4 \\ x_2 = x_4 \\ x_3 = x_4 \\ x_4 = x_4 \end{array} \implies \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 0 & -1 & 1 & 1 \\ -1 & 1 & -2 & 3 \\ 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \lambda\mathbf{x} = (1) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Using $\lambda_3 = 2$ for $(A - \lambda I_4)$ in the augmented matrix $[(A - \lambda I_4) \mathbf{0}]$:

$$\begin{bmatrix} -2 & -1 & 1 & 1 & 0 \\ -1 & -1 & -2 & 3 & 0 \\ 2 & -1 & -2 & 0 & 0 \\ 1 & -1 & 1 & -2 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{matrix} x_1 = x_4 \\ x_2 = 0 \\ x_3 = x_4 \\ x_4 = x_4 \end{matrix} \implies \mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 0 & -1 & 1 & 1 \\ -1 & 1 & -2 & 3 \\ 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}, \quad \lambda\mathbf{x} = (2) \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}$$

Based on the results, we have the following eigenspaces.

$$E_{-1} = \text{span} \left(\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right), \quad E_1 = \text{span} \left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right), \quad E_2 = \text{span} \left(\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right)$$

4.5. Checking if matrices are invertible and/or diagonalizable. Whether a matrix is invertible can be checked easily by calculating the determinant. Whether a matrix is diagonalizable is more difficult. By the Diagonalization Theorem, an $n \times n$ matrix is diagonalizable if, and only if, it has n linearly independent eigenvectors. Another theorem (4.21) says that a symmetric matrix is always diagonalizable.

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

($I D$) This is the 2×2 identity matrix. It is its own inverse and it is on a diagonal form.

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

($\mathcal{I} D$) The determinant is 0 so it is NOT invertible. It is on a diagonal form. (Or: it is symmetric, so it is diagonalizable.)

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

($I \mathcal{D}$) The determinant is 1 so it is invertible. Not showing all the calculations, but it has eigenvalue $\lambda = 1$ with algebraic multiplicity of 2, but it only has one eigenvector. By the Diagonalization Theorem, it must have two linearly independent eigenvectors, so this matrix is NOT diagonalizable.

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

($\mathcal{I} \mathcal{D}$) Eigenvector is 0 so it is NOT invertible. It has eigenvalue $\lambda = 0$ with algebraic multiplicity of 2, but this one also has only one eigenvector. By the Diagonalization Theorem, this matrix is NOT diagonalizable.

4.6. Computing the eigenspace and checking if the matrices are diagonalizable.

a.

$$A = \begin{bmatrix} 2 & 3 & 0 \\ 1 & 4 & 3 \\ 0 & 0 & 1 \end{bmatrix} \implies (A - \lambda I_3) = \begin{bmatrix} 2 - \lambda & 3 & 0 \\ 1 & 4 - \lambda & 3 \\ 0 & 0 & 1 - \lambda \end{bmatrix}$$

Finding the characteristic polynomial (going along the first column).

$$\begin{aligned} p(\lambda) = \det(A - \lambda I_3) &= (2 - \lambda) \begin{vmatrix} 4 - \lambda & 3 \\ 0 & 1 - \lambda \end{vmatrix} - (1) \begin{vmatrix} 3 & 0 \\ 0 & 1 - \lambda \end{vmatrix} \\ &= (2 - \lambda)((4 - \lambda)(1 - \lambda) - 0) - (1)((3)(1 - \lambda) - 0) \\ &= (2 - \lambda)(\lambda^2 - 5\lambda + 4) - (3 - 3\lambda) \\ &= -\lambda^3 + 7\lambda^2 - 11\lambda + 5 \\ &= (\lambda - 1)^2(\lambda - 5) \end{aligned}$$

By inspection $\lambda_1 = 1$ is a root since $p(1) = 0$. After polynomial division and the quadratic formula $\lambda_2 = 5$, where $\lambda_1 = 1$ is a root with algebraic multiplicity of 2.

Calculating the eigenvectors by solving $(A - \lambda I_3)\mathbf{x} = \mathbf{0}$.

Setting in $\lambda = 1$ and solving $[(A - \lambda I_3) \mathbf{0}]$.

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 1 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = -3x_2 \\ x_2 = x_2 \\ x_3 = 0 \end{array} \implies \mathbf{x} = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 2 & 3 & 0 \\ 1 & 4 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \quad \lambda\mathbf{x} = (1) \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}$$

Setting in $\lambda = 5$ and solving $[(A - \lambda I_3) \mathbf{0}]$.

$$\begin{bmatrix} -3 & 3 & 0 & 0 \\ 1 & 1 & 3 & 0 \\ 0 & 0 & -4 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = x_2 \\ x_2 = x_2 \\ x_3 = 0 \end{array} \implies \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 2 & 3 & 0 \\ 1 & 4 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix}, \quad \lambda\mathbf{x} = (5) \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix}$$

There are only two eigenvectors, and though they are linearly independent this matrix is not diagonalizable by the Diagonalization Theorem. The eigenspaces are:

$$E_1 = \text{span} \left(\begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} \right), \quad E_5 = \text{span} \left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right)$$

b.

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies (A - \lambda I_4) = \begin{bmatrix} 1 - \lambda & 1 & 0 & 0 \\ 0 & -\lambda & 0 & 0 \\ 0 & 0 & -\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{bmatrix}$$

Finding the characteristic polynomial, going down the first column.

$$p(\lambda) = \det(A - \lambda I) = (1 - \lambda)(-\lambda^3)$$

By direct inspection we see that the roots are $\lambda_1 = 1$ and $\lambda_2 = 0$ with algebraic multiplicity of 3.

Setting in $\lambda = 1$ and solving $[(A - \lambda I_4) \mathbf{0}]$.

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \sim \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = x_1 \\ x_2 = 0 \\ x_3 = 0 \\ x_4 = 0 \end{array} \implies \mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \lambda\mathbf{x} = (1) \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Setting in $\lambda = 0$ and solving $[(A - \lambda I_4) \mathbf{0}]$.

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = -x_2 \\ x_2 = x_2 \\ x_3 = x_3 \\ x_4 = x_4 \end{array} \implies \mathbf{x} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \lambda\mathbf{x} = (0) \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The two other vectors clearly give $\mathbf{0}$. We have three, linearly independent vectors (by inspection, not calculated), so there are a total of four linearly independent eigenvectors. By the Diagonalization Theorem, A is diagonalizable. The eigenspace E_1 is spanned by the single vector found above, and E_0 is spanned by the three vectors just found.

4.7. Determining if the following matrices are diagonalizable. If they are, we find the diagonal form such that $A = PDP^{-1}$.

a.

$$A = \begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix} \implies (A - \lambda I_2) = \begin{bmatrix} -\lambda & 1 \\ -8 & 4 - \lambda \end{bmatrix}$$

Characteristic polynomial.

$$\det(A - \lambda I_2) = \begin{vmatrix} -\lambda & 1 \\ -8 & 4 - \lambda \end{vmatrix} = \lambda^2 - 4\lambda + 8$$

This characteristic polynomial has no real solutions, only complex solutions. With the quadratic formula we find that $\lambda_1 = 2 - 2i$ and $\lambda_2 = 2 + 2i$.

Using $\lambda_1 = 2 - 2i$

$$\begin{bmatrix} -2 + 2i & 1 & 0 \\ -8 & 2 + 2i & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & -0.25 - 0.25i & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies \begin{matrix} x_1 = (0.25 + 0.25i)x_2 \\ x_2 = x_2 \end{matrix} \implies \mathbf{x} = \begin{bmatrix} 0.25 + 0.25i \\ 1 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix} \begin{bmatrix} 0.25 + 0.25i \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 - 2i \end{bmatrix}, \quad \lambda\mathbf{x} = (2 - 2i) \begin{bmatrix} 0.25 + 0.25i \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 - 2i \end{bmatrix}$$

Using $\lambda_2 = 2 + 2i$

$$\begin{bmatrix} -2 - 2i & 1 & 0 \\ -8 & 2 - 2i & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & -0.25 + 0.25i & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies \begin{matrix} x_1 = (0.25 - 0.25i)x_2 \\ x_2 = x_2 \end{matrix} \implies \mathbf{x} = \begin{bmatrix} 0.25 - 0.25i \\ 1 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix} \begin{bmatrix} 0.25 - 0.25i \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 + 2i \end{bmatrix}, \quad \lambda\mathbf{x} = (2 + 2i) \begin{bmatrix} 0.25 - 0.25i \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 + 2i \end{bmatrix}$$

There are two linearly independent eigenvectors, so by the Diagonalization Theorem this matrix has a diagonal form. From the eigenvectors we can construct the matrix P . We can also calculate the inverse P^{-1} .

$$P = \begin{bmatrix} 0.25 - 0.25i & 0.25 + 0.25i \\ 1 & 1 \end{bmatrix} \implies P^{-1} = \begin{bmatrix} 2i & 0.5 - 0.5i \\ -2i & 0.5 + 0.5i \end{bmatrix}$$

The diagonal matrix D has the eigenvalues on the diagonal (corresponding to the eigenvectors):

$$D = \begin{bmatrix} 2 + 2i & 0 \\ 0 & 2 - 2i \end{bmatrix}$$

By doing all the calculations, we can verify that:

$$PDP^{-1} = \begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix} = A.$$

b.

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \implies A - \lambda I_3 = \begin{bmatrix} 1 - \lambda & 1 & 1 \\ 1 & 1 - \lambda & 1 \\ 1 & 1 & 1 - \lambda \end{bmatrix}$$

First observe that this matrix is *symmetric*, which means it is diagonalizable. The rows are not linearly independent, but all that matters is that the eigenvectors are. Finding the characteristic polynomial. (Skipping the intermediary calculations).

$$p(\lambda) = \det(A - \lambda I_3) = -\lambda^3 + 3\lambda^2 = \lambda^2(3 - \lambda)$$

From this expression we can immediately see that $\lambda_1 = 0$ is a root with algebraic multiplicity of 2, as well as $\lambda_2 = 3$.

Using $\lambda_1 = 0$ for $(A - \lambda I_3)$ in the augmented matrix $[(A - \lambda I_3) \mathbf{0}]$:

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = -x_2 - x_3 \\ x_2 = x_2 \\ x_3 = x_3 \end{array} \implies \mathbf{x} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \lambda\mathbf{x} = (0) \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

It's essentially the same calculation for the other vector. In this case we have both an algebraic multiplicity of 2 and a geometric multiplicity of 2.

Using $\lambda_2 = 3$ for $(A - \lambda I_3)$ in the augmented matrix $[(A - \lambda I_3) \mathbf{0}]$:

$$\begin{bmatrix} -2 & 1 & 1 & 0 \\ 1 & -2 & 1 & 0 \\ 1 & 1 & -2 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = x_3 \\ x_2 = x_3 \\ x_3 = x_3 \end{array} \implies \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}, \quad \lambda\mathbf{x} = (3) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$

We have three linearly independent eigenvectors, so we can again confirm that A is diagonalizable. Constructing the matrix P for eigenvalues 3, 0, 0.

$$P = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \implies P^{-1} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

By calculating we can confirm that $PDP^{-1} = A$.

c.

$$A = \begin{bmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{bmatrix} \implies (A - \lambda I_4) = \begin{bmatrix} 5 - \lambda & 4 & 2 & 1 \\ 0 & 1 - \lambda & -1 & -1 \\ -1 & -1 & 3 - \lambda & 0 \\ 1 & 1 & -1 & 2 - \lambda \end{bmatrix}$$

Characteristic polynomial. (Without calculations).

$$p(\lambda) = \det(A - \lambda I_4) = \lambda^4 - 11\lambda^3 + 42\lambda^2 - 64\lambda + 32$$

This polynomial has the solutions: $\lambda_1 = 4$ (with multiplicity 2), $\lambda_2 = 2$, $\lambda_3 = 1$.

Using $\lambda_1 = 4$ for $(A - \lambda I_4)$ in the augmented matrix $[(A - \lambda I_4) \mathbf{0}]$:

$$\begin{bmatrix} 1 & 4 & 2 & 1 \\ 0 & -4 & -1 & -1 \\ -1 & -1 & -2 & 0 \\ 1 & 1 & -1 & -3 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \implies \begin{matrix} x_1 = 0 \\ x_2 = 0 \\ x_3 = 0 \\ x_4 = 0 \end{matrix} \implies \mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

In order for a 4x4 matrix to be diagonalizable it has to have 4 linearly independent vectors. In this case we only have one vector for a root with algebraic multiplicity of 2. There is also no need to verify that this is an eigenvector since it is obvious that $A\mathbf{0} = \mathbf{0}$. We can check the other two eigenvalues to verify that we only get 3 linearly independent eigenvectors.

Using $\lambda_2 = 2$ for $(A - \lambda I_4)$ in the augmented matrix $[(A - \lambda I_4) \mathbf{0}]$:

$$\begin{bmatrix} 3 & 4 & 2 & 1 \\ 0 & -1 & -1 & -1 \\ -1 & -1 & 1 & 0 \\ 1 & 1 & -1 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{matrix} x_1 = x_4 \\ x_2 = -x_4 \\ x_3 = 0 \\ x_4 = x_4 \end{matrix} \implies \mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 0 \\ 2 \end{bmatrix}, \quad \lambda\mathbf{x} = (2) \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 0 \\ 2 \end{bmatrix}$$

Using $\lambda_3 = 1$ for $(A - \lambda I_4)$ in the augmented matrix $[(A - \lambda I_4) \mathbf{0}]$:

$$\begin{bmatrix} 4 & 4 & 2 & 1 \\ 0 & 0 & -1 & -1 \\ -1 & -1 & 2 & 0 \\ 1 & 1 & -1 & 1 \end{bmatrix} \implies \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{matrix} x_1 = -x_2 \\ x_2 = x_2 \\ x_3 = 0 \\ x_4 = 0 \end{matrix} \implies \mathbf{x} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \lambda\mathbf{x} = (1) \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

As suspected we only have three linearly independent eigenvectors. By the Diagonalization Theorem, this matrix is NOT diagonalizable.

d.

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix} \implies (A - \lambda I_3) = \begin{bmatrix} 5 - \lambda & -6 & -6 \\ -1 & 4 - \lambda & 2 \\ 3 & -6 & -4 - \lambda \end{bmatrix}$$

Characteristic polynomial:

$$p(\lambda) = -\lambda^3 + 5\lambda^2 - 8\lambda + 4 = -(x-2)^2(x-1)$$

Solutions are $\lambda_1 = 2$ and $\lambda_2 = 1$, where the 2 has multiplicity 2. Now we need to check that the corresponding eigenvectors are linearly independent.

Using $\lambda_1 = 2$ for $(A - \lambda I_3)$ in the augmented matrix $[(A - \lambda I_3) \mathbf{0}]$:

$$\begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix} \implies \begin{bmatrix} 1 & -2 & -2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = 2x_2 + 2x_3 \\ x_2 = x_2 \\ x_3 = x_3 \end{array} \implies \mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$

We have two free variables and get two vectors for $\lambda_1 = 2$. Verifying:

$$A\mathbf{x} = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}, \quad \lambda\mathbf{x} = (2) \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$$

And for the second eigenvector.

$$A\mathbf{x} = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix}, \quad \lambda\mathbf{x} = (2) \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix}$$

For both of these eigenvectors we see that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

Using $\lambda_2 = 1$ for $(A - \lambda I_3)$ in the augmented matrix $[(A - \lambda I_3) \mathbf{0}]$:

$$\begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1/3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{matrix} x_1 = x_3 \\ x_2 = -x_3/3 \\ x_3 = x_3 \end{matrix} \implies \mathbf{x} = \begin{bmatrix} 1 \\ -1/3 \\ 1 \end{bmatrix}$$

Verifying:

$$A\mathbf{x} = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix} \begin{bmatrix} 1 \\ -1/3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1/3 \\ 1 \end{bmatrix}, \quad \lambda\mathbf{x} = (1) \begin{bmatrix} 1 \\ -1/3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1/3 \\ 1 \end{bmatrix}$$

We have three linearly independent eigenvectors so this matrix is diagonalizable by the Diagonalization Theorem. The diagonal matrix and the matrix made up of the corresponding eigenvectors:

$$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 0 & -1/3 \\ 0 & 1 & 1 \end{bmatrix} \implies P^{-1} = \begin{bmatrix} -1 & 3 & 2 \\ 3 & -6 & -5 \\ -3 & 6 & 6 \end{bmatrix}$$

Confirming that $PDP^{-1} = A$. Start with PD .

$$\begin{aligned} PD &= \begin{bmatrix} 2 & 2 & 1 \\ 1 & 0 & -1/3 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2(2) + 2(0) + 1(0) & 2(0) + 2(2) + 1(0) & 2(0) + 2(0) + 1(1) \\ 1(2) + 0(0) + (-1/3)(0) & 1(0) + 0(2) + (-1/3)(0) & 1(0) + 0(0) + (-1/3)(1) \\ 0(2) + 1(0) + 1(0) & 0(0) + 1(2) + 1(0) & 0(0) + 1(0) + 1(1) \end{bmatrix} \\ &= \begin{bmatrix} 4 + 0 + 0 & 0 + 4 + 0 & 0 + 0 + 1 \\ 2 + 0 + 0 & 0 + 0 + 0 & 0 + 0 + (-1/3) \\ 0 + 0 + 0 & 0 + 2 + 0 & 0 + 0 + 1 \end{bmatrix} = \begin{bmatrix} 4 & 4 & 1 \\ 2 & 0 & -1/3 \\ 0 & 2 & 1 \end{bmatrix} \end{aligned}$$

Finally, PDP^{-1} .

$$\begin{aligned} PDP^{-1} &= \begin{bmatrix} 4 & 4 & 1 \\ 2 & 0 & -1/3 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & 2 \\ 3 & -6 & -5 \\ -3 & 6 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 4(-1) + 4(3) + 1(-3) & 4(3) + 4(-6) + 1(6) & 4(2) + 4(-5) + 1(6) \\ 2(-1) + 0(3) + (-1/3)(-3) & 2(3) + 0(-6) + (-1/3)(6) & 2(2) + 0(-5) + (-1/3)(6) \\ 0(-1) + 2(3) + 1(-3) & 0(3) + 2(-6) + 1(6) & 0(2) + 2(-5) + 1(6) \end{bmatrix} \\ &= \begin{bmatrix} -4 + 12 + (-3) & 12 + (-24) + 6 & 8 + (-20) + 6 \\ -2 + 0 + 1 & 6 + 0 + (-2) & 4 + 0 + (-2) \\ 0 + 6 + (-3) & 0 + (-12) + 6 & 0 + (-10) + 6 \end{bmatrix} = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix} = A \end{aligned}$$

And we have confirmed that $PDP^{-1} = A$.

4.8. Finding the SVD (singular value decomposition) of:

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

That is, we are going to find a decomposition $U\Sigma V^T = A$. Following the instructions laid out in the text book, there are three steps:

1. **V: Right-singular vectors.** We calculate $A^T A$ which gives us a square matrix. We calculate the eigenvalues and eigenvectors as we do for normal diagonalization, except we only use orthonormal eigenvectors which will make up the columns of V . Remember to TRANSPOSE!
2. **Σ : Singular matrix.** The singular values are the square roots of the eigenvalues found in step 1. Note that the singular matrix must have the same form as the original matrix A with the singular values on the diagonal.
3. **U: Left-singular vectors.** The columns of U are calculated with the equation:

$$\mathbf{u}_i = \frac{1}{\sigma_i} A \mathbf{v}_i.$$

Comment: This method is how we calculate it by hand. It is done differently in a computer (due to certain difficulties with matrix calculations such as poor approximations and rounding errors).

We calculate the matrix $A^T A$ (and call it W for short):

$$W = A^T A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix} \implies (W - \lambda I_3) = \begin{bmatrix} 13 - \lambda & 12 & 2 \\ 12 & 13 - \lambda & -2 \\ 2 & -2 & 8 - \lambda \end{bmatrix}$$

Characteristic polynomial.

$$p(\lambda) = \det(W - \lambda I_3) = -\lambda^3 + 34\lambda^2 - 225\lambda = \lambda(\lambda - 25)(\lambda - 9)$$

The roots are $\lambda_1 = 25$, $\lambda_2 = 9$ and $\lambda_3 = 0$.

Using $\lambda_1 = 25$ for $(W - \lambda I_3)$ in the augmented matrix $[(W - \lambda I_3) \mathbf{0}]$:

$$\begin{bmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{bmatrix} \implies \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = x_2 \\ x_2 = x_2 \\ x_3 = 0 \end{array} \implies \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Using $\lambda_2 = 9$ for $(W - \lambda I_3)$ in the augmented matrix $[(W - \lambda I_3) \mathbf{0}]$:

$$\begin{bmatrix} 4 & 12 & 2 \\ 12 & 4 & -2 \\ 2 & -2 & -1 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & -1/4 & 0 \\ 0 & 1 & 1/4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = 1 \\ x_2 = -1 \\ x_3 = 4 \end{array} \implies \mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}$$

Using $\lambda_3 = 0$ for $(W - \lambda I_3)$ in the augmented matrix $[(W - \lambda I_3) \mathbf{0}]$:

$$\begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{array}{l} x_1 = -2x_3 \\ x_2 = 2x_3 \\ x_3 = x_3 \end{array} \implies \mathbf{x} = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$$

Any symmetric matrix is always diagonalizable, and $W = A^T A$ is symmetric, so these eigenvectors must be linearly independent. Also note that all eigenvalues are squares!

Let us call the eigenvectors \mathbf{x}_1 , \mathbf{x}_2 and \mathbf{x}_3 . We need to translate them into an orthonormal basis. We first check if they are orthogonal:

$$\mathbf{x}_1\mathbf{x}_2 = 0, \mathbf{x}_1\mathbf{x}_3 = 0, \mathbf{x}_2\mathbf{x}_3 = 0$$

Then we make them orthonormal using the dot product. Calculating the length of each vector.

$$\begin{aligned}\|\mathbf{x}_1\| &= \sqrt{1^2 + 1^2 + 0} = \sqrt{2} \\ \|\mathbf{x}_2\| &= \sqrt{1^2 + (-1)^2 + 4^2} = \sqrt{18} = 3\sqrt{2} \\ \|\mathbf{x}_3\| &= \sqrt{(-2)^2 + 2^2 + 1^2} = \sqrt{9} = 3\end{aligned}$$

Creating an orthonormal basis (ONB).

$$\begin{aligned}\mathbf{v}_1 &= \frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \\ \mathbf{v}_2 &= \frac{1}{\|\mathbf{x}_2\|}\mathbf{x}_2 = \begin{bmatrix} 1/3\sqrt{2} \\ -1/3\sqrt{2} \\ 4/3\sqrt{2} \end{bmatrix} \\ \mathbf{v}_3 &= \frac{1}{\|\mathbf{x}_3\|}\mathbf{x}_3 = \begin{bmatrix} -2/3 \\ 2/3 \\ 1/3 \end{bmatrix}\end{aligned}$$

These are the columns of the right-singular matrix.

$$V = \begin{bmatrix} 1/\sqrt{2} & 1/3\sqrt{2} & -2/3 \\ 1/\sqrt{2} & -1/3\sqrt{2} & 2/3 \\ 0 & 4/3\sqrt{2} & 1/3 \end{bmatrix} \implies V^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/3\sqrt{2} & -1/3\sqrt{2} & 4/3\sqrt{2} \\ -2/3 & 2/3 & 1/3 \end{bmatrix}$$

Finding the singular matrix is quite straightforward: square root of the eigenvalues across the diagonal. Since $\text{rank}(A) = 2$ there are only 2 nonzero singular values.

$$\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}$$

All that remains is finding the columns of U . Recall the formula:

$$\mathbf{u}_i = \frac{1}{\sigma_i}A\mathbf{v}_i.$$

Since A is a 2×3 matrix, this means U is a 2×2 matrix, so we only have to calculate it for the nonzero singular values.

$$\begin{aligned}\mathbf{u}_1 &= \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{1}{5} \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 5/\sqrt{2} \\ 5/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \\ \mathbf{u}_2 &= \frac{1}{\sigma_2}A\mathbf{v}_2 = \frac{1}{3} \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} \begin{bmatrix} 1/3\sqrt{2} \\ -1/3\sqrt{2} \\ 4/3\sqrt{2} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 9/3\sqrt{2} \\ -9/3\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}\end{aligned}$$

This gives us the matrix U .

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

Now putting it all together and checking that we get back A :

$$\begin{aligned} U\Sigma V^T &= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/3\sqrt{2} & -1/3\sqrt{2} & 4/3\sqrt{2} \\ -2/3 & 2/3 & 1/3 \end{bmatrix} \\ &= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 5/\sqrt{2} & 5/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 4/\sqrt{2} \end{bmatrix} \\ &= \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} \\ &= A \end{aligned}$$

This confirms that the calculations are correct.

4.9. Finding the singular value decomposition (SVD) of

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

Square matrices that are symmetric and positive-definite have the property that the regular diagonalization and the SVD are equal. This matrix is not symmetric, so we have to do the full calculations. To simplify the notation a bit, we set $W = A^T A$.

$$W = A^T A = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix} \implies (W - \lambda I_2) = \begin{bmatrix} 8 - \lambda & 0 \\ 0 & 2 - \lambda \end{bmatrix}$$

Finding the characteristic polynomial.

$$\det(W - \lambda I_2) = \begin{vmatrix} 8 - \lambda & 0 \\ 0 & 2 - \lambda \end{vmatrix} = (8 - \lambda)(2 - \lambda) \implies \lambda_1 = 8, \lambda_2 = 2$$

Using $\lambda_1 = 8$ for $(W - \lambda I_3)$ in the augmented matrix $[(W - \lambda I_3) \mathbf{0}]$:

$$\begin{bmatrix} 0 & 0 \\ 0 & -6 \end{bmatrix} \implies \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies \begin{matrix} x_1 = x_1 \\ x_2 = 0 \end{matrix} \implies \mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Using $\lambda_2 = 2$ for $(W - \lambda I_3)$ in the augmented matrix $[(W - \lambda I_3) \mathbf{0}]$:

$$\begin{bmatrix} 6 & 0 \\ 0 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \implies \begin{matrix} x_1 = 0 \\ x_2 = x_2 \end{matrix} \implies \mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

These eigenvectors, $\mathbf{x}_1, \mathbf{x}_2$ are orthonormal, so we can set:

$$\begin{aligned}\mathbf{v}_1 &= \mathbf{x}_1 \\ \mathbf{v}_2 &= \mathbf{x}_2\end{aligned}$$

Which gives us the matrix V .

$$V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \implies V^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The singular value matrix contains the square root of the eigenvalues.

$$\Sigma = \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

Finally we calculate the columns of U .

$$\begin{aligned}\mathbf{u}_1 &= \frac{1}{\sigma_1} A \mathbf{v}_1 = \frac{1}{2\sqrt{2}} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{2\sqrt{2}} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ -1/2\sqrt{2} \end{bmatrix} \\ \mathbf{u}_2 &= \frac{1}{\sigma_2} A \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}\end{aligned}$$

From these we construct U .

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

Confirming our calculations.

$$\begin{aligned}U \Sigma V^T &= \begin{bmatrix} 1/\sqrt{2} & 2/\sqrt{2} \\ -1/2\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1/\sqrt{2} & 2/\sqrt{2} \\ -1/2\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \\ &= \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \\ &= A\end{aligned}$$

4.10. Finding the rank-1 approximation.

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

As explained in the text book, we can find the rank-1 matrix with the following formula:

$$A_1 = \mathbf{u}_1 \mathbf{v}_1^T.$$

The terminology is a little confusing. We are interested in the rank- k approximation for $k = 1$ which is this:

$$\hat{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T.$$

A is the same matrix as we used in exercise 4.8, so we can borrow the vectors we found in order to do the calculations.

$$\begin{aligned} \hat{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T &= (5) \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} [1/\sqrt{2} \quad 1/\sqrt{2} \quad 0] \\ &= (5) \begin{bmatrix} (1/\sqrt{2})(1/\sqrt{2}) & (1/\sqrt{2})(1/\sqrt{2}) & (1/\sqrt{2})(0) \\ (1/\sqrt{2})(1/\sqrt{2}) & ((1/\sqrt{2})1/\sqrt{2}) & (1/\sqrt{2})(0) \end{bmatrix} \\ &= (5) \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2.5 & 2.5 & 0 \\ 2.5 & 2.5 & 0 \end{bmatrix} \end{aligned}$$

This matrix is a very crude approximation of A .

4.11.

Before moving on to the main exercise, we will prove a lemma that we will need.

Lemma. For a matrix $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A^T A) = \text{rank}(A)$.

Proof.

Assuming we have some matrix $A \in \mathbb{R}^{m \times n}$. First we will prove that $A^T A$ and A have the same kernel.

⊂) Assume $\mathbf{x} \in \ker(A)$. Then:

$$\begin{aligned} A\mathbf{x} &= \mathbf{0} \\ A^T A\mathbf{x} &= A^T \mathbf{0} \\ A^T A\mathbf{x} &= \mathbf{0} \end{aligned}$$

Since $A^T A\mathbf{x} = \mathbf{0}$ this means that $\mathbf{x} \in \ker(A^T A)$ which means $\ker(A) \subset \ker(A^T A)$.

⊃) Assume now that $\mathbf{x} \in \ker(A^T A)$. Then:

$$\begin{aligned} A^T A\mathbf{x} &= \mathbf{0} \\ \mathbf{x}^T A^T A\mathbf{x} &= \mathbf{x}^T \mathbf{0} \\ \mathbf{x}^T A^T A\mathbf{x} &= 0 \\ (A\mathbf{x})^T A\mathbf{x} &= 0 \\ \|A\mathbf{x}\|^2 &= 0 \quad (\dagger - \text{see notes}) \\ A\mathbf{x} &= \mathbf{0} \end{aligned}$$

Since $A\mathbf{x} = \mathbf{0}$ this means that $\mathbf{x} \in \ker(A)$ which means $\ker(A^T A) \subset \ker(A)$. By inclusion both ways, we can conclude that $\ker(A) = \ker(A^T A)$.

Both A and $A^T A$ map into \mathbb{R}^n . (Columns are input and rows correspond to image/range, which is \mathbb{R}^n for both matrices). Let's say that $\dim(\ker(A)) = k$ for some $k \leq \min(m, n)$ and by what we just showed, $\dim(\ker(A^T A)) = k$. By the rank-nullity theorem:

$$\begin{aligned} \dim(\mathbb{R}^n) &= n = \text{rank}(A) + \dim(\ker(A)) = \text{rank}(A) + k \\ \dim(\mathbb{R}^n) &= n = \text{rank}(A^T A) + \dim(\ker(A^T A)) = \text{rank}(A^T A) + k \end{aligned}$$

So:

$$\text{rank}(A) + k = \text{rank}(A^T A) + k \implies \text{rank}(A) = \text{rank}(A^T A)$$

which shows that $\text{rank}(A) = \text{rank}(A^T A)$. ■

Regarding (†). We multiply a $m \times n$ matrix with a $n \times 1$ vector, we end up with a $m \times 1$ vector, so:

$$A\mathbf{x} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}, \quad (A\mathbf{x})^T = [a_1 \quad \cdots \quad a_m].$$

When we multiply them:

$$(A\mathbf{x})^T A\mathbf{x} = [a_1 \quad \cdots \quad a_m] \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = (a_1)^2 + \cdots + (a_m)^2 = \|A\mathbf{x}\|^2.$$

Since all the elements are squared and added, it can only be $\mathbf{0}$ when all values are 0, hence $A\mathbf{x} = \mathbf{0}$.

Corollary. The rank of a symmetric matrix A equals the number of nonzero eigenvalues.

Proof. Will not prove this, can be found in various linear algebra text books. Maybe later...

Now we go to the main exercise.

Claim: For any $A \in \mathbb{R}^{m \times n}$ the matrices $A^T A$ and AA^T possess the same nonzero eigenvalues.

Proof.

Assume A is an $m \times n$ matrix with $\text{rank}(A) = k$ for some $k \in \mathbb{N}$ and $1 \leq k \leq \min(m, n)$. From A we can create the matrices $A^T A$ and AA^T which are both symmetric (i.e. it is its own transpose):

$$\begin{aligned} (A^T A)^T &= A^T (A^T)^T = A^T A \\ (AA^T)^T &= (A^T)^T A^T = AA^T \end{aligned}$$

By the Lemma we proved above (we can set $B := A^T$ and use the Lemma on B and $B^T B = AA^T$):

$$\text{rank}(A) = k = \text{rank}(A^T A) = \text{rank}(AA^T).$$

For our matrices $A^T A$ has rank of k and therefore k nonzero eigenvalues, by the unproven corollary. The same is true for AA^T . In other words $A^T A$ and AA^T have the same number of nonzero eigenvalues.

Finally we must show that they have the same eigenvalues. Assume $\lambda \neq 0$ is an eigenvalue for $A^T A$ with the corresponding eigenvector $\mathbf{x} \in \mathbb{R}^n$. Define $\mathbb{R}^m \ni \mathbf{y} := A\mathbf{x}$. We have:

$$\begin{aligned} A^T A\mathbf{x} &= \lambda\mathbf{x} \\ AA^T A\mathbf{x} &= A\lambda\mathbf{x} = \lambda A\mathbf{x} \\ AA^T \mathbf{y} &= \lambda\mathbf{y} \end{aligned}$$

Which shows that λ is an eigenvalue for AA^T with corresponding eigenvector \mathbf{y} . We can also show the reverse for some eigenvalue $\gamma \neq 0$ for AA^T with eigenvector $\mathbf{x} \in \mathbb{R}^m$. Define $\mathbb{R}^n \ni \mathbf{y} := A^T \mathbf{x}$.

$$\begin{aligned} AA^T \mathbf{x} &= \gamma\mathbf{x} \\ A^T AA^T \mathbf{x} &= A^T \gamma\mathbf{x} = \gamma A^T \mathbf{x} \\ A^T A\mathbf{y} &= \gamma\mathbf{y} \end{aligned}$$

which means γ is an eigenvalue for $A^T A$. In conclusion we have shown that AA^T and $A^T A$ have the same nonzero eigenvalues. ■

4.12.

Showing that for $\mathbf{x} \neq \mathbf{0}$, Theorem 4.24 holds. That is,

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

where σ_1 is the largest singular value of $A \in \mathbb{R}^{m \times n}$. (Not using the $\|\cdot\|_2$ notation.)

Proof.

This can be verified by using the SVD and making some clever observations. By the SVD we can write $A = U\Sigma V^T$. Since U and V are orthonormal matrices, they have the property that $V^T = V^{-1}$, so $U^T U = I$ and $V V^T = I$. (Also using that $(V^T)^T = V$ without mention).

For reasons that will become clear, we will define $\mathbf{y} := V^T \mathbf{x}$. Note then that:

$$\|\mathbf{y}\|^2 = \mathbf{y}^T \mathbf{y} = (V^T \mathbf{x})^T V^T \mathbf{x} = \mathbf{x}^T V V^T \mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|^2.$$

We will also use the following:

$$\begin{aligned} A^T A &= (U\Sigma V^T)^T U\Sigma V^T \\ &= V\Sigma^T U^T U\Sigma V^T \\ &= V\Sigma^T \Sigma V^T \\ &= V S V^T, \end{aligned} \tag{♣}$$

where we define S to be the symmetric $n \times n$ matrix (empty spaces are 0s):

$$S = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix}$$

The dimensions are $n \times n$ because A is an $m \times n$ matrix, which means Σ is an $m \times n$ matrix, Σ^T is an $n \times m$ matrix and $\Sigma^T \Sigma$ becomes an $n \times n$ matrix.

Observe now that:

$$\begin{aligned} \|A\mathbf{x}\|^2 &= (A\mathbf{x})^T A\mathbf{x} \\ &= \mathbf{x}^T A^T A\mathbf{x} \\ &= \mathbf{x}^T V S V^T \mathbf{x} \\ &= \mathbf{y}^T S \mathbf{y} \end{aligned} \tag{♣}$$

(Continues on next page.)

$$\begin{aligned}
&= [y_1 \quad \dots \quad y_n] \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \\
&= [y_1 \quad \dots \quad y_n] \begin{bmatrix} \sigma_1^2 y_1 \\ \vdots \\ \sigma_n^2 y_n \end{bmatrix} \\
&= \sigma_1^2 y_1^2 + \dots + \sigma_n^2 y_n^2 \\
&\leq \sigma_1^2 y_1^2 + \dots + \sigma_1^2 y_n^2 && \text{(Replacing all } \sigma_i^2 \text{ with } \sigma_1^2\text{)} \\
&= \sigma_1^2 (y_1^2 + \dots + y_n^2) \\
&= \sigma_1^2 \|\mathbf{y}\|^2 \\
&= \sigma_1^2 \|\mathbf{x}\|^2
\end{aligned}$$

By these calculations, we have shown that:

$$\|A\mathbf{x}\|^2 \leq \sigma_1^2 \|\mathbf{x}\|^2 \implies \|A\mathbf{x}\| \leq \sigma_1 \|\mathbf{x}\| \implies \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \leq \sigma_1$$

We can avoid the inequality by selecting:

$$\mathbf{x} = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} \implies \mathbf{y} = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix},$$

which specifically targets σ_1 and which then gives us equality. Hence we have shown that:

$$\max_{\mathbf{x}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \sigma_1 \quad \blacksquare$$

Chapter 5: Vector Calculus - Exercises

5.1. Computing the derivative for

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

This is a combination of the product rule, $(uv)' = u'v + uv'$, and the chain rule.

$$\begin{aligned} f'(x) &= \left(\frac{1}{x^4}\right) 4x^3 \sin(x^3) + \log(x^4) \cos(x^3) 3x^2 \\ &= \frac{4 \sin(x^3)}{x} + 3x^2 \log(x^4) \cos(x^3) \end{aligned}$$

5.2. Computing the derivative of the logistic sigmoid.

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

This is differentiating with the quotient rule:

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}.$$

Finding the derivative:

$$\begin{aligned} f'(x) &= \frac{(0)(1 + e^{-x}) - (1)(-e^{-x})}{(1 + e^{-x})^2} \\ &= \frac{e^{-x}}{1 + 2e^{-x} + e^{-2x}} \end{aligned}$$

5.3. Finding the derivative of a function similar to the Gaussian pdf. Rewriting the expression a bit. This is the chain rule twice.

$$f(x) = \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

$$\begin{aligned} f'(x) &= \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) \left(-\frac{1}{\sigma^2}(x - \mu)\right) \\ &= \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) \left(-\frac{1}{\sigma^2}(x - \mu)\right) \end{aligned}$$

Alternative form:

$$f'(x) = -\frac{(x - \mu) \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)}{\sigma^2}$$

Another alternative:

$$f'(x) = \frac{\mu - x}{\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) = \frac{\mu - x}{\sigma^2} f(x)$$

5.4 Computing the Taylor polynomials T_n for $n = 0, \dots, 5$ at $x_0 = 0$ for the function:

$$f(x) = \sin(x) + \cos(x).$$

Recalling the general Taylor polynomial form (Definition 5.3):

$$T_n(x) := \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k.$$

• $n = 0$.

$$T_0(x) = f(0)(x)^0 = (\sin(0) + \cos(0))x = 1$$

• $n = 1$.

$$\begin{aligned} T_1(x) &= T_0 + f'(0)(x)^1 \\ &= T_0(x) + (\cos(0) - \sin(0))x \\ &= 1 + x \end{aligned}$$

• $n = 2$.

$$\begin{aligned} T_2(x) &= T_1(x) + \frac{f''(0)}{2!}(x)^2 \\ &= T_1(x) + \frac{(-\sin(0) - \cos(0))}{2}x^2 \\ &= 1 + x - \frac{x^2}{2} \end{aligned}$$

• $n = 3$.

$$\begin{aligned} T_3(x) &= T_2(x) + \frac{f'''(0)}{3!}(x)^3 \\ &= T_2(x) + \frac{(-\cos(0) + \sin(0))}{6}x^3 \\ &= 1 + x - \frac{x^2}{2} - \frac{x^3}{6} \end{aligned}$$

• $n = 4$.

$$\begin{aligned} T_4(x) &= T_3(x) + \frac{f^{(4)}(0)}{4!}(x)^4 \\ &= T_3(x) + \frac{(\sin(0) + \cos(0))}{24}x^4 \\ &= 1 + x - \frac{x^2}{2} - \frac{x^3}{6} + \frac{x^4}{24} \end{aligned}$$

• $n = 5$.

$$\begin{aligned} T_5(x) &= T_4(x) + \frac{f^{(5)}(0)}{5!}(x)^5 \\ &= T_4(x) + \frac{(\cos(0) - \sin(0))}{120}x^5 \\ &= 1 + x - \frac{x^2}{2} - \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} \end{aligned}$$

5.5. For each function we will determine:

- (a) The dimension of $\frac{\partial f}{\partial \mathbf{x}}$.
- (b) The Jacobian.

Function:

$$f_1(\mathbf{x}) = \sin(x_1) \cos(x_2)$$

a.

Writing the derivative in vector form:

$$\frac{\partial f_1}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \end{bmatrix}$$

so the dimension of $\frac{\partial f_1}{\partial \mathbf{x}}$ is (1×2) since there is one function f_1 and two variables.

b.

Before calculating the Jacobian, we differentiate the functions f_1 wrt. each variable.

$$\begin{aligned} \frac{\partial}{\partial x_1} \sin(x_1) \cos(x_2) &= \cos(x_1) \cos(x_2) \\ \frac{\partial}{\partial x_2} \sin(x_1) \cos(x_2) &= -\sin(x_1) \sin(x_2) \end{aligned}$$

Collecting the terms in a matrix, as per Definition 5.6 (or Example 5.7).

$$\mathbf{J} = \begin{bmatrix} \cos(x_1) \cos(x_2) & -\sin(x_1) \sin(x_2) \end{bmatrix}$$

Function:

$$f_2(\mathbf{x}) = \mathbf{x}^T \mathbf{y} = x_1 y_1 + \dots + x_n y_n$$

where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

a.

Writing the derivative in vector form:

$$\frac{\partial f_2}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_2}{\partial x_n} \end{bmatrix}.$$

In this case we also have one function f_2 , but n variables, so the dimension is $(1 \times n)$.

b.

Differentiating the functions f_2 wrt. each variable x_i .

$$\begin{aligned} \frac{\partial}{\partial x_1} (x_1 y_1 + \dots + x_n y_n) &= y_1 \\ &\vdots \\ \frac{\partial}{\partial x_n} (x_1 y_1 + \dots + x_n y_n) &= y_n \end{aligned}$$

Collecting the terms in a matrix.

$$\mathbf{J} = [y_1 \dots y_n]$$

Function:

$$f_3(\mathbf{x}) = \mathbf{xx}^T = \begin{bmatrix} x_1^2 & x_1x_2 & \dots & x_1x_n \\ x_1x_2 & x_2^2 & \dots & x_2x_n \\ \vdots & & \ddots & \\ x_1x_n & x_2x_n & \dots & x_n^2 \end{bmatrix}$$

where $\mathbf{x} \in \mathbb{R}^n$.

a.

Writing the derivative:

$$\frac{\partial f_3}{\partial \mathbf{x}} = \left\{ \frac{\partial f_3}{\partial x_i} = \begin{bmatrix} \frac{\partial}{\partial x_i} x_1^2 & \frac{\partial}{\partial x_i} x_1x_2 & \dots & \frac{\partial}{\partial x_i} x_1x_n \\ \frac{\partial}{\partial x_i} x_1x_2 & \frac{\partial}{\partial x_i} x_2^2 & \dots & \frac{\partial}{\partial x_i} x_2x_n \\ \vdots & & \ddots & \\ \frac{\partial}{\partial x_i} x_1x_n & \frac{\partial}{\partial x_i} x_2x_n & \dots & \frac{\partial}{\partial x_i} x_n^2 \end{bmatrix} : 1 \leq i \leq n \right\}.$$

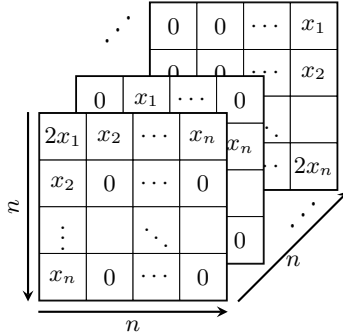
For every x_i in \mathbf{x} , we differentiate the $(n \times n)$ matrix wrt. x_i and create a tensor gradient, with dimension $(n \times n \times n)$.

b.

Differentiating wrt. each variable x_i to get the Jacobian tensor which is illustrated below. Each matrix in the tensor will contain the calculated derivatives listed above. For instance, for x_2 we get the following matrix - which is the second layer in the illustration:

$$\frac{\partial f_3}{\partial x_2} = \begin{bmatrix} 0 & x_1 & \dots & 0 \\ x_1 & 2x^2 & \dots & x_n \\ \vdots & & \ddots & \\ 0 & x_n & \dots & 0 \end{bmatrix}.$$

Illustration of the Jacobian tensor, where we have included $\frac{\partial f_3}{\partial x_1}$, $\frac{\partial f_3}{\partial x_2}$ and $\frac{\partial f_3}{\partial x_n}$:



To make the last exercise a bit easier to understand, we will solve it again for a much simpler example. Take the same function, but with $\mathbf{x} \in \mathbb{R}^2$:

$$f(\mathbf{x}) = \mathbf{x}\mathbf{x}^T = \begin{bmatrix} x_1^2 & x_1x_2 \\ x_1x_2 & x_2^2 \end{bmatrix}$$

Now, the derivative becomes just two matrices:

$$\frac{df}{d\mathbf{x}} = \left\{ \frac{\partial f}{\partial x_1} = \begin{bmatrix} \frac{\partial}{\partial x_1} x_1^2 & \frac{\partial}{\partial x_1} x_1x_2 \\ \frac{\partial}{\partial x_1} x_1x_2 & \frac{\partial}{\partial x_1} x_2^2 \end{bmatrix} = \begin{bmatrix} 2x_1 & x_2 \\ x_2 & 0 \end{bmatrix}, \frac{\partial f}{\partial x_2} = \begin{bmatrix} \frac{\partial}{\partial x_2} x_1^2 & \frac{\partial}{\partial x_2} x_1x_2 \\ \frac{\partial}{\partial x_2} x_1x_2 & \frac{\partial}{\partial x_2} x_2^2 \end{bmatrix} = \begin{bmatrix} 0 & x_1 \\ x_1 & 2x_2 \end{bmatrix} \right\}$$

These will also make a tensor, but only a $(2 \times 2 \times 2)$ tensor.

$$\text{Layer 1} = \begin{bmatrix} 2x_1 & x_2 \\ x_2 & 0 \end{bmatrix}$$

$$\text{Layer 2} = \begin{bmatrix} 0 & x_1 \\ x_1 & 2x_2 \end{bmatrix}$$

5.6. Differentiating the function f wrt. \mathbf{t} and g wrt. to \mathbf{X} where:

$$f(\mathbf{t}) = \sin(\log(\mathbf{t}^T \mathbf{t})), \quad \mathbf{t} \in \mathbb{R}^D$$

$$g(\mathbf{X}) = \text{tr}(\mathbf{A}\mathbf{X}\mathbf{B}), \quad \mathbf{A} \in \mathbb{R}^{D \times E}, \mathbf{X} \in \mathbb{R}^{E \times F}, \mathbf{B} \in \mathbb{R}^{F \times D},$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix (sum of the diagonal elements).

Since $\mathbf{t} \in \mathbb{R}^D$:

$$\mathbf{t}^T \mathbf{t} = t_1^2 + t_2^2 + \dots + t_D^2.$$

Differentiating f will give us a vector:

$$\frac{df}{d\mathbf{t}} = \left[\frac{\partial f}{\partial t_1} \quad \frac{\partial f}{\partial t_2} \quad \dots \quad \frac{\partial f}{\partial t_D} \right]$$

Calculating for t_1 with the chain rule:

$$\frac{\partial f}{\partial t_1} = \frac{\partial}{\partial t_1} (\sin(\log(t_1^2 + t_2^2 + \dots + t_D^2))) = \cos(\log(\mathbf{t}^T \mathbf{t})) \cdot \frac{1}{\mathbf{t}^T \mathbf{t}} 2t_1 = \frac{2t_1 \cos(\log(\mathbf{t}^T \mathbf{t}))}{\mathbf{t}^T \mathbf{t}}$$

Similarly, for $1 \leq i \leq D$:

$$\frac{\partial f}{\partial t_i} = \frac{2t_i \cos(\log(\mathbf{t}^T \mathbf{t}))}{\mathbf{t}^T \mathbf{t}}$$

Which results in:

$$\frac{df}{d\mathbf{t}} = \left[\frac{2t_1 \cos(\log(\mathbf{t}^T \mathbf{t}))}{\mathbf{t}^T \mathbf{t}} \quad \frac{2t_2 \cos(\log(\mathbf{t}^T \mathbf{t}))}{\mathbf{t}^T \mathbf{t}} \quad \dots \quad \frac{2t_D \cos(\log(\mathbf{t}^T \mathbf{t}))}{\mathbf{t}^T \mathbf{t}} \right]$$

We have the matrix $\mathbf{A} \in \mathbb{R}^{D \times E}$, $\mathbf{X} \in \mathbb{R}^{E \times F}$ and $\mathbf{B} \in \mathbb{R}^{F \times D}$, where:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1F} \\ x_{21} & x_{22} & \dots & x_{2F} \\ \vdots & & \ddots & \\ x_{E1} & x_{E2} & \dots & x_{EF} \end{bmatrix}.$$

By matrix multiplication, $\mathbf{AXB} \in \mathbb{R}^{D \times D}$, so it is a square matrix which is required for the trace. If we define the diagonal elements to be d_i for $1 \leq i \leq D$. Then the trace will give:

$$\text{trace}(\mathbf{AXB}) = \sum_{i=1}^D d_i \in \mathbb{R}$$

Before we continue, let us do the calculations for a simplified case. Let's set:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

By doing the matrix multiplications and summing the diagonal elements (see **Appendix** for the calculations), we find that:

$$\begin{aligned} \text{trace}(\mathbf{AXB}) &= a_{11}b_{11}x_{11} + a_{11}b_{21}x_{12} + a_{12}b_{11}x_{21} + a_{12}b_{21}x_{22} \\ &\quad + a_{21}b_{12}x_{11} + a_{21}b_{22}x_{12} + a_{22}b_{12}x_{21} + a_{22}b_{22}x_{22} \end{aligned}$$

Since the trace is a real number with dimension (1×1) , we know that $\frac{dg}{d\mathbf{X}}$ will have dimensions $(1 \times 1) \times (2 \times 2)$, which just becomes a 2×2 matrix. Now, we write out all the derivatives we need for the gradient.

$$\frac{dg}{d\mathbf{X}} = \left\{ \frac{\partial \text{trace}(\mathbf{AXB})}{\partial x_{ij}} : 1 \leq i \leq 2, 1 \leq j \leq 2 \right\}$$

Calculating the individual partial derivatives.

$$\begin{aligned} \frac{\partial \text{trace}(\mathbf{AXB})}{\partial x_{11}} &= a_{11}b_{11} + a_{21}b_{12} \\ \frac{\partial \text{trace}(\mathbf{AXB})}{\partial x_{12}} &= a_{11}b_{21} + a_{21}b_{22} \\ \frac{\partial \text{trace}(\mathbf{AXB})}{\partial x_{21}} &= a_{12}b_{11} + a_{22}b_{12} \\ \frac{\partial \text{trace}(\mathbf{AXB})}{\partial x_{22}} &= a_{12}b_{21} + a_{22}b_{22} \end{aligned}$$

Which means:

$$\begin{aligned} \frac{dg}{d\mathbf{X}} &= \begin{bmatrix} a_{11}b_{11} + a_{21}b_{12} & a_{11}b_{21} + a_{21}b_{22} \\ a_{12}b_{11} + a_{22}b_{12} & a_{12}b_{21} + a_{22}b_{22} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{bmatrix} \\ &= \mathbf{A}^T \mathbf{B}^T \end{aligned} \tag{†}$$

This works since all matrices are squares. But this still works in the general case since $\mathbf{A}^T \in \mathbb{R}^{E \times D}$ and $\mathbf{B}^T \in \mathbb{R}^{D \times F}$ so $\mathbf{A}^T \mathbf{B}^T \in \mathbb{R}^{E \times F}$ which will have the same dimension as \mathbf{X} .

Returning to the main exercise. We need to find a complete expression for the trace, which is a bit involved, but also some good repetition of some linear algebra theory.

Let us start by defining $\mathbf{W} := \mathbf{X}\mathbf{B}$. From standard matrix multiplication we can write out the expression for w_{ij} the element of \mathbf{W} on row i and column j . It is:

$$w_{ij} = \sum_{k=1}^F x_{ik} b_{kj}. \quad (\diamond)$$

(Which also shows that w_{ij} is the inner product of row i from \mathbf{X} and column j from \mathbf{B}).

Now we define $\mathbf{P} := \mathbf{A}\mathbf{W} = \mathbf{A}(\mathbf{X}\mathbf{B})$. By standard matrix multiplication:

$$\begin{aligned} p_{ij} &= \sum_{k=1}^E a_{ik} w_{kj} \\ &= \sum_{k=1}^E a_{ik} \sum_{k'=1}^F x_{kk'} b_{k'j} \\ &= \sum_{k=1}^E \sum_{k'=1}^F a_{ik} x_{kk'} b_{k'j} \end{aligned} \quad (\diamond)$$

(We can move the term a_{ik} in to the sum because it is independent of the index k'). Recall that we need the trace, which is all elements on the diagonal of \mathbf{P} , or in other words all elements where $i = j$. We are also going to sum these elements up, so we get an explicit expression for the trace like this:

$$\text{trace}(\mathbf{A}\mathbf{X}\mathbf{B}) = \sum_{d=1}^D p_{dd} = \sum_{d=1}^D \sum_{k=1}^E \sum_{k'=1}^F a_{dk} x_{kk'} b_{k'd}$$

(By setting $D = 2$, $E = 2$ and $F = 2$, we can see that this expression recreates the trace from the simplified example.) Since the trace is a real number with dimension (1×1) , we know that $\frac{dg}{d\mathbf{X}}$ will have dimensions $(1 \times 1) \times (E \times F)$, which just becomes an $E \times F$ matrix. Now, we write out all the derivatives we need for the gradient.

$$\frac{dg}{d\mathbf{X}} = \left\{ \frac{\partial \text{trace}(\mathbf{A}\mathbf{X}\mathbf{B})}{\partial x_{ij}} : 1 \leq i \leq E, 1 \leq j \leq F \right\}$$

Next we calculate the individual derivatives. Whenever the x_{ij} we are differentiating on is NOT in the term, the derivative becomes 0, which eliminates a lot of the terms in the sums. First when differentiating wrt. x_{11} , only the terms where $k = 1$ and $k' = 1$ are kept since all others become 0. Then we can remove two of the sums:

$$\frac{\partial \text{trace}(\mathbf{A}\mathbf{X}\mathbf{B})}{\partial x_{11}} = \sum_{d=1}^D a_{d1} b_{1d}$$

And in the more general case when differentiating wrt. x_{ij} :

$$\frac{\partial \text{trace}(\mathbf{A}\mathbf{X}\mathbf{B})}{\partial x_{ij}} = \sum_{d=1}^D a_{di} b_{jd}$$

When comparing this to (\diamond) we see that the rows and columns for i, j have switched, and recalling the simplified example in (\ddagger) we conclude that $\frac{dg}{d\mathbf{X}} = \mathbf{A}^T \mathbf{B}^T \in \mathbb{R}^{E \times F}$.

5.7.

Finding the derivatives $df/d\mathbf{x}$ by using the chain rule. Also describing the dimension of all partial derivatives.

a.

The function:

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

Finding z expressed in x_i .

$$z = \mathbf{x}^T \mathbf{x} = x_1^2 + x_2^2 + \dots + x_D^2$$

This derivative can be calculated by using the chain rule like this:

$$\frac{df}{d\mathbf{x}} = \frac{\partial f}{\partial z} \frac{\partial z}{\partial \mathbf{x}}$$

In this case: $\partial f/\partial z$ has dimension (1×1) and $\partial z/\partial \mathbf{x}$ has dimension $(1 \times D)$. Calculating the derivatives.

$$\frac{\partial f}{\partial z} = \frac{\partial}{\partial z} (\log(1 + z)) = \frac{1}{1 + z}$$

Calculating the derivative for a specific x_i .

$$\frac{\partial z}{\partial x_i} = \frac{\partial}{\partial x_i} (x_1^2 + x_2^2 + \dots + x_D^2) = 2x_i$$

When differentiating over the vector \mathbf{x} we have the same pattern.

$$\frac{\partial z}{\partial \mathbf{x}} = [2x_1 \quad 2x_2 \quad \dots \quad 2x_D]$$

Putting it all together:

$$\frac{df}{d\mathbf{x}} = \frac{\partial f}{\partial z} \frac{\partial z}{\partial \mathbf{x}} = \frac{1}{1 + z} [2x_1 \quad 2x_2 \quad \dots \quad 2x_D] = \left[\frac{2x_1}{1 + z} \quad \frac{2x_2}{1 + z} \quad \dots \quad \frac{2x_D}{1 + z} \right]$$

b.

The function:

$$f(\mathbf{z}) = \sin(\mathbf{z}), \quad \mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

where $\mathbf{A} \in \mathbb{R}^{E \times D}$, $\mathbf{x} \in \mathbb{R}^D$, $\mathbf{b} \in \mathbb{R}^E$. From the matrix multiplication we see that $\mathbf{z} \in \mathbb{R}^E$, and as stated the sinus function is used on all elements in the column vector. So:

$$f(\mathbf{z}) = \sin(\mathbf{z}) = \begin{bmatrix} \sin(\mathbf{z}) \\ \sin(\mathbf{z}) \\ \vdots \\ \sin(\mathbf{z}) \end{bmatrix} = \begin{bmatrix} \sin(z_1) \\ \sin(z_2) \\ \vdots \\ \sin(z_E) \end{bmatrix}.$$

As in equation (5.54) we have a vector valued function. We have to calculate the derivative in each row for all z_i as shown in (5.56). We keep the E rows and get E columns of partial derivatives. From this we can see that $\partial f/\partial \mathbf{z}$ has dimensions $(E \times E)$.

Writing out the expression for \mathbf{z} explicitly.

$$\mathbf{z} = \mathbf{Ax} + \mathbf{b}$$

$$\begin{aligned}
 &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1D} \\ a_{21} & a_{22} & \cdots & a_{2D} \\ \vdots & & \ddots & \\ a_{E1} & a_{E2} & \cdots & a_{ED} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_E \end{bmatrix} \\
 &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1D}x_D \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2D}x_D \\ \vdots \\ a_{E1}x_1 + a_{E2}x_2 + \cdots + a_{ED}x_D \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_E \end{bmatrix} \\
 &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1D}x_D + b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2D}x_D + b_2 \\ \vdots \\ a_{E1}x_1 + a_{E2}x_2 + \cdots + a_{ED}x_D + b_E \end{bmatrix} \\
 &= \begin{bmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_E(\mathbf{x}) \end{bmatrix}
 \end{aligned}$$

Since $\mathbf{z} \in \mathbb{R}^D$ and $\mathbf{x} \in \mathbb{R}^E$, the derivative $\partial \mathbf{z} / \partial \mathbf{x}$ will have dimensions $(E \times D)$ (we keep the E columns, but add one column per derivative of x_i).

By the chain rule:

$$\frac{df}{d\mathbf{x}} = \frac{\partial f}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}}$$

Calculating the derivatives:

$$\frac{\partial f}{\partial \mathbf{z}} = \begin{bmatrix} \frac{\partial}{\partial z_1} \sin(z_1) & \frac{\partial}{\partial z_2} \sin(z_1) & \cdots & \frac{\partial}{\partial z_E} \sin(z_1) \\ \frac{\partial}{\partial z_1} \sin(z_2) & \frac{\partial}{\partial z_2} \sin(z_2) & \cdots & \frac{\partial}{\partial z_E} \sin(z_2) \\ \vdots & & \ddots & \\ \frac{\partial}{\partial z_1} \sin(z_E) & \frac{\partial}{\partial z_2} \sin(z_E) & \cdots & \frac{\partial}{\partial z_E} \sin(z_E) \end{bmatrix} = \begin{bmatrix} \cos(z_1) & 0 & \cdots & 0 \\ 0 & \cos(z_2) & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \cos(z_E) \end{bmatrix}$$

For the next derivative, let us consider a specific case:

$$\frac{\partial z_2}{\partial x_3} = \frac{\partial g_2}{\partial x_3} = \frac{\partial}{\partial x_3} (a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2D}x_D + b_2) = a_{23}$$

This lets us see the general pattern:

$$\frac{\partial z_i}{\partial x_j} = \frac{\partial g_i}{\partial x_j} = a_{ij}$$

The full derivative becomes, when using the temporary $g(\cdot)$ we used earlier:

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_D} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_D} \\ \vdots & & \ddots & \\ \frac{\partial g_E}{\partial x_1} & \frac{\partial g_E}{\partial x_2} & \cdots & \frac{\partial g_E}{\partial x_D} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1D} \\ a_{21} & a_{22} & \cdots & a_{2D} \\ \cdots & & \ddots & \\ a_{E1} & a_{E2} & \cdots & a_{ED} \end{bmatrix} = \mathbf{A}$$

Putting it all together:

$$\frac{df}{d\mathbf{x}} = \frac{\partial f}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \begin{bmatrix} \cos(z_1) & 0 & \cdots & 0 \\ 0 & \cos(z_2) & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \cos(z_E) \end{bmatrix} \mathbf{A}$$

5.8. Computing the derivatives $df/d\mathbf{x}$, and providing dimensions of partial derivatives.

a.

The function:

$$\begin{aligned} f(z) &= \exp\left(-\frac{1}{2}z\right) \\ z &= g(\mathbf{y}) = \mathbf{y}^T \mathbf{S}^{-1} \mathbf{y} \\ \mathbf{y} &= h(\mathbf{x}) = \mathbf{x} - \boldsymbol{\mu} \end{aligned}$$

where $\mathbf{y}, \mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^D$ and $\mathbf{S} \in \mathbb{R}^{D \times D}$. Also note that $z \in \mathbb{R}$.

Writing out \mathbf{y} explicitly.

$$\mathbf{y} = \mathbf{x} - \boldsymbol{\mu} = \begin{bmatrix} h_1(\mathbf{x}) \\ h_2(\mathbf{x}) \\ \vdots \\ h_D(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \\ x_D - \mu_D \end{bmatrix}$$

Writing out z explicitly. (Note that we use the s_{ii} for the inverted matrix \mathbf{S}^{-1}).

$$\begin{aligned}
z &= \mathbf{y}^T \mathbf{S}^{-1} \mathbf{y} \\
&= [y_1 \ y_2 \ \dots \ y_D] \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1D} \\ s_{21} & s_{22} & \dots & s_{2D} \\ \vdots & & \ddots & \\ s_{D1} & s_{D2} & \dots & s_{DD} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_D \end{bmatrix} \\
&= [y_1 \ y_2 \ \dots \ y_D] \begin{bmatrix} s_{11}y_1 + s_{12}y_2 + \dots + s_{1D}y_D \\ s_{21}y_1 + s_{22}y_2 + \dots + s_{2D}y_D \\ \vdots \\ s_{D1}y_1 + s_{D2}y_2 + \dots + s_{DD}y_D \end{bmatrix} \\
&= s_{11}y_1^2 + s_{12}y_1y_2 + \dots + s_{1D}y_1y_D \\
&\quad + s_{21}y_1y_2 + s_{22}y_2^2 + \dots + s_{2D}y_2y_D + \dots \\
&\quad + s_{D1}y_1y_D + s_{D2}y_2y_D + \dots + s_{DD}y_D^2
\end{aligned}$$

We can calculate the derivate with the following chain:

$$\frac{df}{d\mathbf{x}} = \frac{\partial f}{\partial z} \frac{\partial z}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}}$$

Finding the partial derivatives and their dimension.

$$\frac{\partial f}{\partial z} = \frac{\partial}{\partial z} \left(\exp \left(-\frac{1}{2}z \right) \right) = -\frac{1}{2} \exp \left(-\frac{1}{2}z \right)$$

Since z is a real number, this has (1×1) dimensions.

Now to calculate $\partial z / \partial \mathbf{y}$. We differentiate z wrt. all y_i so it will become a row vector with dimension $(1 \times D)$.

$$\frac{\partial z}{\partial \mathbf{y}} = \left[\frac{\partial z}{\partial y_1} \quad \frac{\partial z}{\partial y_2} \quad \dots \quad \frac{\partial z}{\partial y_D} \right]$$

Let us first differentiate for a specific value y_1 to see the pattern.

$$\begin{aligned}
\frac{\partial z}{\partial y_1} &= 2s_{11}y_1 + s_{12}y_2 + \dots + s_{1D}y_D + s_{21}y_2 + \dots + s_{D1}y_D \\
&= 2s_{11}y_1 + y_2(s_{12} + s_{21}) + \dots + y_D(s_{1D} + s_{D1}) \\
&= \sum_{k=1}^D y_k (s_{1k} + s_{k1})
\end{aligned}$$

So the diagonal value gets a 2, and we keep row 1 and column 1. The general pattern is:

$$\frac{\partial z}{\partial y_i} = \sum_{k=1}^D y_k (s_{ik} + s_{ki})$$

Writing out the full derivative.

$$\frac{\partial z}{\partial \mathbf{y}} = \left[\sum_{k=1}^D y_k (s_{1k} + s_{k1}) \quad \dots \quad \sum_{k=1}^D y_k (s_{ik} + s_{ki}) \quad \dots \quad \sum_{k=1}^D y_k (s_{Dk} + s_{kD}) \right]$$

Finally, the last partial derivative, $\partial \mathbf{y} / \partial \mathbf{x}$. We differentiate all values in vector $\mathbf{y} \in \mathbb{R}^D$ by all values in $\mathbf{x} \in \mathbb{R}^D$ so we will get a $(D \times D)$ matrix. Starting with a specific value.

$$\frac{\partial h_1}{\partial x_1} = \frac{\partial}{\partial x_1} (x_1 + \mu_1) = 1$$

And in the first row, we will only get 0 in all other columns. Now, the full matrix.

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \dots & \frac{\partial h_1}{\partial x_D} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \dots & \frac{\partial h_2}{\partial x_D} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_D}{\partial x_1} & \frac{\partial h_D}{\partial x_2} & \dots & \frac{\partial h_D}{\partial x_D} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = I_D$$

This becomes the identity matrix. Putting all of this together (and using that the last one is the identity matrix, and disappears):

$$\begin{aligned} \frac{df}{d\mathbf{x}} &= \frac{\partial f}{\partial z} \frac{\partial z}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \\ &= \frac{\partial f}{\partial z} \frac{\partial z}{\partial \mathbf{y}} \\ &= -\frac{1}{2} \exp\left(-\frac{1}{2}z\right) \left[\sum_{k=1}^D y_k (s_{1k} + s_{k1}) \quad \dots \quad \sum_{k=1}^D y_k (s_{Dk} + s_{kD}) \right] \end{aligned}$$

b.

Function:

$$f(\mathbf{x}) = \text{trace}(\mathbf{x}\mathbf{x}^T + \sigma^2 I_D)$$

where $\mathbf{x} \in \mathbb{R}^D$. The trace is the sum of the diagonals.

First we write out $\mathbf{x}\mathbf{x}^T$ explicitly, which is a $(D \times D)$ matrix.

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \dots & x_D \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1 x_2 & \dots & x_1 x_D \\ x_1 x_2 & x_2^2 & \dots & x_1 x_D \\ \vdots & \vdots & \ddots & \vdots \\ x_1 x_D & x_2 x_D & \dots & x_D^2 \end{bmatrix}$$

Next, we add the other matrix.

$$\mathbf{xx}^T + \sigma^2 I_D = \begin{bmatrix} x_1^2 + \sigma^2 & x_1 x_2 & \dots & x_1 x_D \\ x_1 x_2 & x_2^2 + \sigma^2 & \dots & x_2 x_D \\ \vdots & & & \\ x_1 x_D & x_2 x_D & \dots & x_D^2 + \sigma^2 \end{bmatrix}$$

When applying the trace, we get the following:

$$\begin{aligned} \text{trace}(\mathbf{xx}^T + \sigma^2 I_D) &= x_1^2 + \sigma^2 + x_2^2 + \sigma^2 + \dots + x_D^2 + \sigma^2 \\ &= x_1^2 + x_2^2 + \dots + x_D^2 + D\sigma^2 \end{aligned}$$

When finding the derivative, there are no variables we can use the chain rule on, so we differentiate \mathbf{x} directly. Since the trace is a real number, and we differentiate it on every x_i we get a $(1 \times D)$ vector with partial derivatives. Differentiating on one specific value:

$$\frac{\partial f}{\partial x_2} = \frac{\partial}{\partial x_2} (x_1^2 + x_2^2 + \dots + x_D^2 + D\sigma^2) = 2x_2$$

The complete derivative:

$$\frac{df}{d\mathbf{x}} = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \dots \quad \frac{\partial f}{\partial x_D} \right] = \left[2x_1 \quad 2x_2 \quad \dots \quad 2x_D \right] = 2\mathbf{x}^T$$

c.

In this case we don't have to calculate the product of the partial derivatives explicitly. The function:

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

We have $\mathbf{z} \in \mathbb{R}^M$. Chain of partial derivatives:

$$\frac{df}{d\mathbf{x}} = \frac{\partial f}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}}$$

Notice that $\partial \mathbf{z} / \partial \mathbf{x}$ is the same as in exercise 5.7(b), where we showed:

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \mathbf{A}.$$

with dimensions $(M \times N)$. (The calculations are the same, we just have to change the dimensions).

For the first partial derivative (we introduce f_i to make later calculations a bit easier):

$$f(\mathbf{z}) = \tanh(\mathbf{z}) = \begin{bmatrix} \tanh(\mathbf{z}) \\ \tanh(\mathbf{z}) \\ \vdots \\ \tanh(\mathbf{z}) \end{bmatrix} = \begin{bmatrix} \tanh(z_1) \\ \tanh(z_2) \\ \vdots \\ \tanh(z_M) \end{bmatrix} = \begin{bmatrix} f_1(z_1) \\ f_2(z_2) \\ \vdots \\ f_M(z_M) \end{bmatrix}$$

We differentiate the vector on all z_i which will give us a $(M \times M)$ matrix. Differentiating on one specific value:

$$\frac{\partial}{\partial z_1} (\tanh(z_1)) = \text{sech}^2(z_1)$$

which will be 0 when differentiating on z_2, \dots, z_M . Writing out the full derivative:

$$\frac{\partial f}{\partial \mathbf{z}} = \begin{bmatrix} \frac{\partial f_1}{\partial z_1} & \frac{\partial f_1}{\partial z_2} & \dots & \frac{\partial f_1}{\partial z_M} \\ \frac{\partial f_2}{\partial z_1} & \frac{\partial f_2}{\partial z_2} & \dots & \frac{\partial f_2}{\partial z_M} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial z_1} & \frac{\partial f_M}{\partial z_2} & \dots & \frac{\partial f_M}{\partial z_M} \end{bmatrix} = \begin{bmatrix} \text{sech}^2(z_1) & 0 & \dots & 0 \\ 0 & \text{sech}^2(z_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{sech}^2(z_M) \end{bmatrix}$$

The full derivative:

$$\frac{df}{d\mathbf{x}} = \frac{\partial f}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}}$$

$$= \begin{bmatrix} \text{sech}^2(z_1) & 0 & \dots & 0 \\ 0 & \text{sech}^2(z_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \text{sech}^2(z_M) \end{bmatrix} \mathbf{A}$$

5.9. We have the functions:

$$g(\mathbf{z}, \boldsymbol{\nu}) = \log p(\mathbf{x}, \mathbf{z}) - \log q(\mathbf{z}, \boldsymbol{\nu})$$

$$\mathbf{z} = t(\boldsymbol{\epsilon}, \boldsymbol{\nu})$$

for differentiable functions p, q, t , and $\mathbf{x} \in \mathbb{R}^D$, $\mathbf{z} \in \mathbb{R}^E$, $\boldsymbol{\nu} \in \mathbb{R}^F$, $\boldsymbol{\epsilon} \in \mathbb{R}^G$. With the chain rule, we will compute the gradient

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

Replacing \mathbf{z} with the function t . The full expression becomes:

$$g(\mathbf{z}, \boldsymbol{\nu}) = \log p(\mathbf{x}, t(\boldsymbol{\epsilon}, \boldsymbol{\nu})) - \log q(t(\boldsymbol{\epsilon}, \boldsymbol{\nu}), \boldsymbol{\nu})$$

Let's also set $f(w) = \log(w)$. Using that we can split the derivative over subtraction, we get the following chain:

$$\frac{d}{d\boldsymbol{\nu}} g(\mathbf{z}, \boldsymbol{\nu}) = \frac{\partial f}{\partial p} \frac{\partial p}{\partial t} \frac{\partial t}{\partial \boldsymbol{\nu}} - \frac{\partial f}{\partial q} \frac{\partial q}{\partial t} \frac{\partial t}{\partial \boldsymbol{\nu}}$$

We can't really go more into depth since we don't have any more details on the functions.

Chapter 6: Probability and Distributions - Exercises

6.1. We have the following bivariate distribution:

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_1	x_2	x_3	x_4	x_5

a.

Calculating the marginal distribution. First for Y . This amounts to summing up the probabilities for each y_i values, ignoring X .

$$P(Y = y_1) = 0.01 + 0.02 + 0.03 + 0.1 + 0.1 = 0.26$$

$$P(Y = y_2) = 0.05 + 0.1 + 0.05 + 0.07 + 0.2 = 0.47$$

$$P(Y = y_3) = 0.1 + 0.05 + 0.03 + 0.05 + 0.04 = 0.27$$

Note that $0.26 + 0.47 + 0.27 = 1$.

Now for X .

$$P(X = x_1) = 0.01 + 0.05 + 0.1 = 0.16$$

$$P(X = x_2) = 0.02 + 0.1 + 0.05 = 0.17$$

$$P(X = x_3) = 0.03 + 0.05 + 0.03 = 0.11$$

$$P(X = x_4) = 0.1 + 0.07 + 0.05 = 0.22$$

$$P(X = x_5) = 0.1 + 0.2 + 0.04 = 0.34$$

Note that $0.16 + 0.17 + 0.11 + 0.22 + 0.34 = 1$.

b.

Calculating the conditional probabilities $P(x|Y = y_1)$ and $P(y|X = x_3)$. Starting with $P(y|X = x_3)$:

$$P(Y = y_1|X = x_3) = 0.03/0.11 \approx 0.273$$

$$P(Y = y_2|X = x_3) = 0.05/0.11 \approx 0.454$$

$$P(Y = y_3|X = x_3) = 0.03/0.11 \approx 0.273$$

These sum up to 1. Now for X .

$$P(X = x_1|Y = y_1) = 0.01/0.26 \approx 0.036$$

$$P(X = x_2|Y = y_1) = 0.02/0.26 \approx 0.077$$

$$P(X = x_3|Y = y_1) = 0.03/0.26 \approx 0.115$$

$$P(X = x_4|Y = y_1) = 0.1/0.26 \approx 0.385$$

$$P(X = x_5|Y = y_1) = 0.1/0.26 \approx 0.385$$

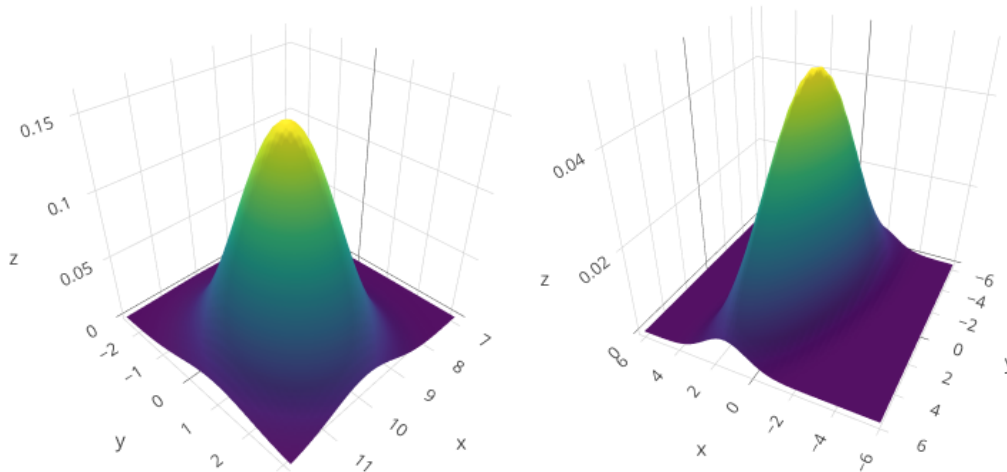
These sum up to 1.

6.2. We have a mixture of two Gaussian distributions.

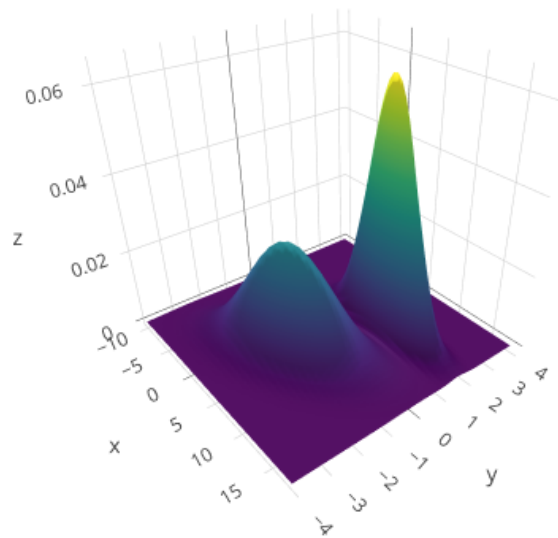
$$0.4\mathcal{N}\left(\begin{bmatrix} 10 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) + 0.6\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 8.4 & 2.0 \\ 2.0 & 1.7 \end{bmatrix}\right)$$

a.

Computing the marginal distribution for each dimension. This question is closely related to example 6.14. Making some plots to investigate the distributions. (Code used for plotting can be found in the **Appendix**). Here are the individual plots:



Note that the maximum values appear at (10, 1) and (0, 0) for each plot. The first plot shows no pattern between x and y which is because the variables are not correlated. This is different in the second plot where it is 'stretched out'. When we combine them with the formula above, we get:



For a joint Gaussian distribution:

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}\right),$$

the marginal distribution is also Gaussian and can be computed by applying the sum rule:

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y}) d\mathbf{y} = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}). \quad (6.68)$$

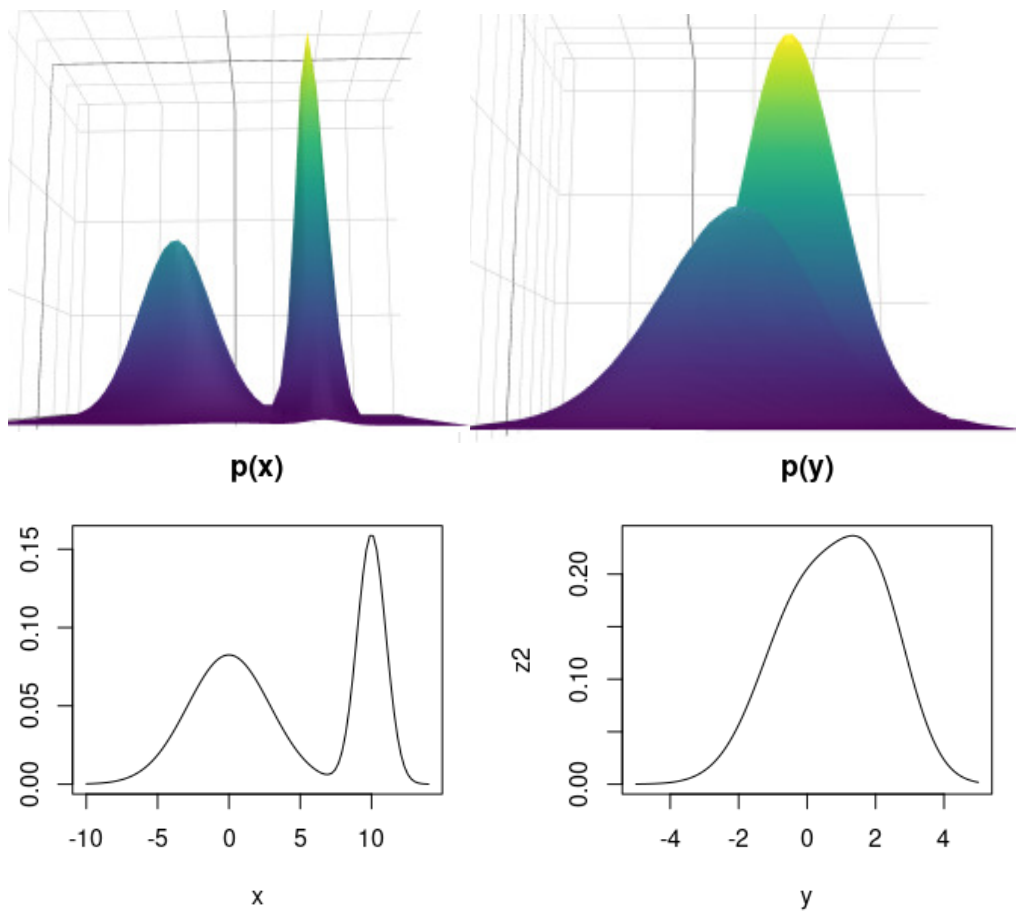
By applying this rule for \mathbf{x} we get that $p(x)$ is the sum of two univariate Gaussian distributions:

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

The same can be done for y (same argument, only with the parameters switched):

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

Here is a comparison of the marginal distributions and the 3d plot viewed from the sides.



b.

The marginal Gaussian distributions above are each the sum of two univariate Gaussian distributions. To find the mean, we can use **Theorem 6.12** from the text book, which tells us that the mean of a weighted sum of Gaussian distributions, is simply the weighted sum of their means.

$$\mu_x = 0.4\mu_1 + 0.6\mu_2 = 0.4(10) + 0.6(0) = 4$$

$$\mu_y = 0.4\mu_1 + 0.6\mu_2 = 0.4(2) + 0.6(0) = 0.8$$

We are asked to find the mean, median and mode for each marginal distribution, and not dimension. One understanding is that this means for each specific Gaussian, which becomes very simple since mean = mode = median for Gaussian distributions.

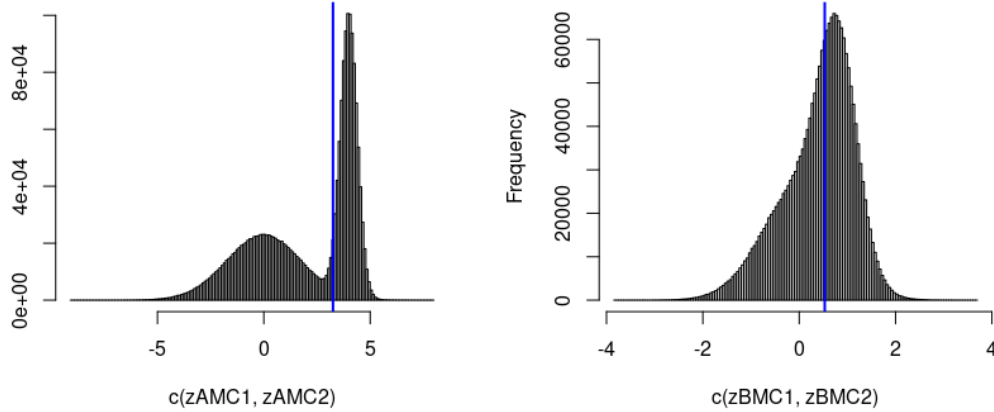
$$\text{Distribution 1 : } \Rightarrow \mu_{x1} = \text{mode}_{x1} = \text{median}_{x1} = 10$$

$$\text{Distribution 2 : } \Rightarrow \mu_{x2} = \text{mode}_{x2} = \text{median}_{x2} = 0$$

$$\text{Distribution 3 : } \Rightarrow \mu_{y1} = \text{mode}_{y1} = \text{median}_{y1} = 2$$

$$\text{Distribution 4 : } \Rightarrow \mu_{y2} = \text{mode}_{y2} = \text{median}_{y2} = 0$$

Another, and more realistic, interpretation is per dimension. This is not possible to solve analytically, so we must solve it numerically. One possibility is with Monte Carlo simulation. Simulating each marginal distribution and marking the median in blue:



Based on this, the values are:

$$p(x) : \Rightarrow \text{median}_x = 3.2519$$

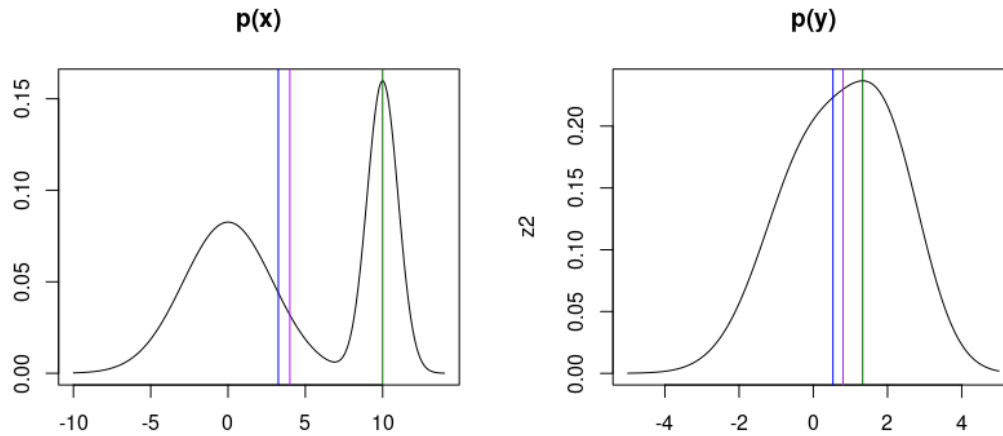
$$p(y) : \Rightarrow \text{median}_y = 0.53023$$

And we can find the mode (only finding the top one for the multimodel distribution):

$$p(x) : \Rightarrow \text{mode}_x = 9.9988$$

$$p(y) : \Rightarrow \text{mode}_y = 1.330633$$

Showing a plot of the distribution with all values highlighted. We are using purple for mean, blue for median, and green for the mode.



c.

To calculate the mean, we use a generalized version of **Theorem 6.12**.

$$\mu_{tot} = 0.4\mu_1 + 0.6\mu_2 = (0.4) \begin{bmatrix} 10 \\ 2 \end{bmatrix} + (0.6)\mathbf{0} = \begin{bmatrix} 4 \\ 0.8 \end{bmatrix}$$

For the mode we have to solve it numerically. We just find what coordinates make the maximum SURF variable, which turns out to be approximately $(10, 1)$.

6.3. We are modelling the success/failure rate of a computer to compile some code by using a Bernoulli distribution, with some parameter $\mu \in (0, 1)$:

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

As described in table **6.2** on page 209 in the text book, we select the Beta distribution as the conjugate prior to the Bernoulli distribution, which will give us a Beta posterior. The Beta distribution has the following pdf on the parameter μ :

$$p(\mu|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \mu^{\alpha-1}(1-\mu)^{\beta-1} \propto \mu^{\alpha-1}(1-\mu)^{\beta-1}$$

Following the calculations from example **6.12**, we can solve a simplified case of the problem.

$$\begin{aligned} p(\mu|x) &= p(x|\mu)p(\mu|\alpha, \beta) \\ &\propto \mu^x(1-\mu)^{1-x} \mu^{\alpha-1}(1-\mu)^{\beta-1} \\ &= \mu^{x+\alpha-1}(1-\mu)^{\beta+(1-x)-1} \\ &\propto \text{Beta}(x + \alpha, \beta + (1 - x)) \end{aligned}$$

But this is just for a single, specific trial x . We want to find the posterior distribution for N trials: x_1, \dots, x_N . For the following calculations we will define:

$$Z = \sum_{i=1}^N x_i$$

(which will become the number of compiling successes). Given μ we can write:

$$p(x_1, \dots, x_N | \mu) \propto \mu^Z (1 - \mu)^{N-Z}.$$

With these simple alterations, we can do the full calculation.

$$\begin{aligned} p(\mu | x_1, \dots, x_N) &= p(x_1, \dots, x_N | \mu) p(\mu | \alpha, \beta) \\ &\propto \mu^Z (1 - \mu)^{N-Z} \mu^{\alpha-1} (1 - \mu)^{\beta-1} \\ &= \mu^{Z+\alpha-1} (1 - \mu)^{\beta+(N-Z)-1} \\ &\propto \text{Beta}(Z + \alpha, \beta + (N - Z)) \end{aligned}$$

6.4. Solving the question by using Baye's formula.

First of all, let us register the coin toss probabilities:

$$P(H) = \frac{3}{5}, \quad P(T) = \frac{2}{5}$$

In bag 1 (H) there are 4 mangos and 2 apples. In bag 2 (T) there are 4 mangos and 4 apples.

We are going to calculate the probability that we used bag 2 if we get a mango. In other words, the evidence is a mango, so what is the probability that we chose bag 2, or that we got a T on the coin toss. In even more precise terms, we are going to calculate $P(T|M)$. From Baye's formula:

$$P(T|M) = \frac{P(M|T)P(T)}{P(M)}$$

We are missing two of these. The first one is $P(M|T)$. This is the probability of getting a mango if we are looking into bag 2, and from the given information this is $2/3$. The other probability, $P(M|H)$ is $1/2$.

For calculating $P(M)$, we use the total law of probability (where $P(H) = P(-T)$):

$$\begin{aligned} P(M) &= P(M|T)P(T) + P(M|-T)P(-T) \\ &= P(M|T)P(T) + P(M|H)P(H) \\ &= \left(\frac{1}{2}\right) \left(\frac{2}{5}\right) + \left(\frac{2}{3}\right) \left(\frac{3}{5}\right) \\ &= \frac{1}{5} + \frac{2}{5} \\ &= \frac{3}{5} \end{aligned}$$

Now we have everything we need:

$$P(T|M) = \frac{P(M|T)P(T)}{P(M)} = \frac{\frac{1}{2} \cdot \frac{2}{5}}{\frac{3}{5}} = \frac{1}{3}$$

6.5. Skipped, for now. I don't remember much about time series. :)

6.6. Proving a pretty standard identity.

$$\begin{aligned}
 V_X[x] &= \mathbb{E}[(x - \mu)^2] \\
 &= \mathbb{E}_X[x^2 - 2x\mu + \mu^2] \\
 &= \mathbb{E}_X[x^2] - 2\mu\mathbb{E}_X[x] + \mu^2 \\
 &= \mathbb{E}_X[x^2] - 2\mu^2 + \mu^2 \\
 &= \mathbb{E}_X[x^2] - \mu^2 \\
 &= \mathbb{E}_X[x^2] - (\mathbb{E}_X[x])^2
 \end{aligned} \tag{6.44}$$

6.7 Showing a second identity. Will use the following formula:

$$\bar{x} = \frac{1}{N} \sum_{k=1}^N x_k \implies \sum_{k=1}^N x_k = N\bar{x}$$

$$\begin{aligned}
 \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (x_i - x_j)^2 &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (x_i^2 - 2x_i x_j + x_j^2) \\
 &= \frac{1}{N^2} \sum_{i=1}^N \left(\sum_{j=1}^N x_i^2 - 2x_i \sum_{j=1}^N x_j + \sum_{j=1}^N x_j^2 \right) \\
 &= \frac{1}{N^2} \sum_{i=1}^N \left(N x_i^2 - 2x_i N\bar{x} + \sum_{j=1}^N x_j^2 \right) \\
 &= \frac{1}{N^2} \left(N \sum_{i=1}^N x_i^2 - 2N\bar{x} \sum_{i=1}^N x_i + \sum_{i=1}^N \sum_{j=1}^N x_j^2 \right) \\
 &= \frac{1}{N^2} \left(N \sum_{i=1}^N x_i^2 - 2N\bar{x} \sum_{i=1}^N x_i + N \sum_{j=1}^N x_j^2 \right) \\
 &= \frac{1}{N} \left(\sum_{i=1}^N x_i^2 + \sum_{j=1}^N x_j^2 - 2\bar{x} \sum_{i=1}^N x_i \right) \\
 &= \frac{1}{N} \left(2 \sum_{i=1}^N x_i^2 - 2 \frac{1}{N} \sum_{i=1}^N x_i \sum_{i=1}^N x_i \right) \\
 &= 2 \left(\frac{1}{N} \sum_{i=1}^N x_i^2 - \frac{1}{N} \sum_{i=1}^N x_i \frac{1}{N} \sum_{i=1}^N x_i \right) \\
 &= 2 \left[\frac{1}{N} \sum_{i=1}^N x_i^2 - \left(\frac{1}{N} \sum_{i=1}^N x_i \right)^2 \right]
 \end{aligned} \tag{6.45}$$

6.8. This is already done in example (6.14). Here is the translation from the standard Bernoulli pdf to the general exponential form:

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

Where $h(x) = 1$, $\phi(x) = x$, and $\theta = \frac{\mu}{1-\mu}$. Using this θ , we can show:

$$\theta = \frac{\mu}{1-\mu} \implies \mu = \frac{1}{1 + \exp(-\theta)}$$

(which is the sigmoid function). Then, using this, we can show:

$$A(\theta) = -\log(1-\mu) = \log(1 + \exp(\theta)).$$

6.9. Doing the same calculations for the Binomial distribution, where the calculations are very similar to exercise 6.8. Translating from the pdf.

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

Where we have the following parameters (by using some of the same arguments as above):

$$h(x) = \binom{N}{x}, \quad \phi(x) = x, \quad \theta = \log \left(\frac{\mu}{1-\mu} \right), \quad A(\theta) = N \log(1 + \exp(\theta))$$

Next, we will find the general exponential form for the Beta distribution. To simplify the notation, we simply set:

$$\Gamma(\alpha, \beta) := \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$\begin{aligned} p(x|\alpha, \beta) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \\ &= \exp \left[\log \left(\Gamma(\alpha, \beta) x^{\alpha-1} (1-x)^{\beta-1} \right) \right] \\ &= \exp \left[(\alpha - 1) \log(x) + (\beta - 1) \log(1-x) - (-\log \Gamma(\alpha, \beta)) \right] \\ &= h(x) \exp \left[\phi(\mathbf{x})^T \boldsymbol{\theta} - A(\boldsymbol{\theta}) \right] \end{aligned}$$

This is the exponential form, with:

$$h(x) = 1, \quad \phi(\mathbf{x}) = \begin{bmatrix} \log(x) \\ \log(1-x) \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \alpha - 1 \\ \beta - 1 \end{bmatrix}, \quad A(\boldsymbol{\theta}) = -\log \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

The last step is to multiply the exponential form of the Binomial and Beta distributions and show that the product is still on the exponential form. We multiply the terms and collect the terms in a suitable way.

$$\begin{aligned} &\binom{N}{x} \exp \left[x \log \left(\frac{\mu}{1-\mu} \right) - (-N \log(1-\mu)) \right] \exp \left[(\alpha - 1) \log(x) + (\beta - 1) \log(1-x) - (-\log \Gamma(\alpha, \beta)) \right] \\ &= \binom{N}{x} \exp \left[x \log \left(\frac{\mu}{1-\mu} \right) - (-N \log(1-\mu)) + (\alpha - 1) \log(x) + (\beta - 1) \log(1-x) - (-\log \Gamma(\alpha, \beta)) \right] \\ &= \binom{N}{x} \exp \left[x \log \left(\frac{\mu}{1-\mu} \right) + (\alpha - 1) \log(x) + (\beta - 1) \log(1-x) - (-\log \Gamma(\alpha, \beta) - N \log(1-\mu)) \right] \\ &= h(x) \exp \left[\phi(\mathbf{x})^T \boldsymbol{\theta} - A(\boldsymbol{\theta}) \right] \end{aligned}$$

Where:

$$h(x) = \binom{N}{x}, \quad \phi(\mathbf{x}) = \begin{bmatrix} x \\ \log(x) \\ \log(1-x) \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \log \left(\frac{\mu}{1-\mu} \right) \\ \alpha - 1 \\ \beta - 1 \end{bmatrix}, \quad A(\boldsymbol{\theta}) = -\log \Gamma(\alpha, \beta) - N \log(1-\mu)$$

The general exponential form is still there, showing that this product is still in the exponential family.

6.10. This result is important for when Bayesian methods are used in e.g. linear regression. When the likelihood and prior are assumed to be Gaussian, they need to be multiplied.

The product of two Gaussian distributions $\mathcal{N}(\mathbf{a}, \mathbf{A})$ and $\mathcal{N}(\mathbf{b}, \mathbf{B})$ is an unnormalized Gaussian distribution $c\mathcal{N}(\mathbf{c}, \mathbf{C})$, where

$$\begin{aligned}\mathbf{C} &= (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} \\ \mathbf{c} &= \mathbf{C}(\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b}) \\ c &= (2\pi)^{-D/2} |\mathbf{A} + \mathbf{B}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{a} - \mathbf{b})^T (\mathbf{A} + \mathbf{B})^{-1} (\mathbf{a} - \mathbf{b})\right)\end{aligned}$$

a.

Solving the multiplication by 'completing the square'. First we write out the full form of the multidimensional Gaussian distributions. Assuming $\mathbf{x}, \mathbf{a}, \mathbf{b} \in \mathbb{R}^D$ and $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{D \times D}$. In this calculation, $|\cdot|$ is the determinant.

$$\begin{aligned}p(\mathbf{x}|\mathbf{a}, \mathbf{A}) &= (2\pi)^{-\frac{D}{2}} |\mathbf{A}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{a})^T \mathbf{A}^{-1} (\mathbf{x} - \mathbf{a})\right) \\ p(\mathbf{x}|\mathbf{b}, \mathbf{B}) &= (2\pi)^{-\frac{D}{2}} |\mathbf{B}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{b})^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{b})\right)\end{aligned}$$

This calculation is VERY extensive, so we have to split it into smaller parts. When multiplying, we get this:

$$p(\mathbf{x}|\mathbf{a}, \mathbf{A})p(\mathbf{x}|\mathbf{b}, \mathbf{B}) = N_1 \exp(W_1) N_2 \exp(W_2) = N_1 N_2 \exp(W_1 + W_2)$$

We will look at each of these separately. Beginning with the exponent $W_1 + W_2$. Written out fully and adding the exponents together:

$$\begin{aligned}W_1 + W_2 &= -\frac{1}{2}(\mathbf{x} - \mathbf{a})^T \mathbf{A}^{-1} (\mathbf{x} - \mathbf{a}) - \frac{1}{2}(\mathbf{x} - \mathbf{b})^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{b}) \\ &= -\frac{1}{2} \left[(\mathbf{x} - \mathbf{a})^T \mathbf{A}^{-1} (\mathbf{x} - \mathbf{a}) + (\mathbf{x} - \mathbf{b})^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{b}) \right] \\ &= -\frac{1}{2} \left[\mathbf{x}^T \mathbf{A}^{-1} \mathbf{x} - \mathbf{x}^T \mathbf{A}^{-1} \mathbf{a} - \mathbf{a}^T \mathbf{A}^{-1} \mathbf{x} + \mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \right. \\ &\quad \left. \mathbf{x}^T \mathbf{B}^{-1} \mathbf{x} - \mathbf{x}^T \mathbf{B}^{-1} \mathbf{b} - \mathbf{b}^T \mathbf{B}^{-1} \mathbf{x} + \mathbf{b}^T \mathbf{B}^{-1} \mathbf{b} \right] \\ &= -\frac{1}{2} \left[\mathbf{x}^T (\mathbf{A}^{-1} + \mathbf{B}^{-1}) \mathbf{x} - \mathbf{x}^T (\mathbf{A}^{-1} \mathbf{a} + \mathbf{B}^{-1} \mathbf{b}) - \right. \\ &\quad \left. (\mathbf{a}^T \mathbf{A}^{-1} + \mathbf{b}^T \mathbf{B}^{-1}) \mathbf{x} + \mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \mathbf{b}^T \mathbf{B}^{-1} \mathbf{b} \right] \tag{U}\end{aligned}$$

Before we continue, we can make some changes.

$$\mathbf{C} = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} \implies \mathbf{C}^{-1} = \mathbf{A}^{-1} + \mathbf{B}^{-1}$$

Writing out the vector \mathbf{c} fully:

$$\mathbf{c} = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} (\mathbf{A}^{-1} \mathbf{a} + \mathbf{B}^{-1} \mathbf{b})$$

Remember that the matrices are symmetric, so their sum is also symmetric, and the transpose does not change the matrices, nor their inverse. Which means we can write the transpose of \mathbf{c} as:

$$\mathbf{c}^T = (\mathbf{a}^T \mathbf{A}^{-1} + \mathbf{b}^T \mathbf{B}^{-1}) (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}$$

With this in mind, we can make these clever replacements (I is the identity matrix):

$$\begin{aligned}
\mathbf{x}^T (\mathbf{A}^{-1} \mathbf{a} + \mathbf{B}^{-1} \mathbf{b}) &= \mathbf{x}^T I (\mathbf{A}^{-1} \mathbf{a} + \mathbf{B}^{-1} \mathbf{b}) \\
&= \mathbf{x}^T \underbrace{(\mathbf{A}^{-1} + \mathbf{B}^{-1})}_{\mathbf{C}^{-1}} \underbrace{(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}}_{\mathbf{C}} (\mathbf{A}^{-1} \mathbf{a} + \mathbf{B}^{-1} \mathbf{b}) \\
&= \mathbf{x}^T \underbrace{(\mathbf{A}^{-1} + \mathbf{B}^{-1})}_{\mathbf{C}^{-1}} \underbrace{(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} (\mathbf{A}^{-1} \mathbf{a} + \mathbf{B}^{-1} \mathbf{b})}_{\mathbf{c}} \\
&= \mathbf{x}^T \mathbf{C}^{-1} \mathbf{c} \\
\\
(\mathbf{a}^T \mathbf{A}^{-1} + \mathbf{b}^T \mathbf{B}^{-1}) \mathbf{x} &= (\mathbf{a}^T \mathbf{A}^{-1} + \mathbf{b}^T \mathbf{B}^{-1}) I \mathbf{x} \\
&= (\mathbf{a}^T \mathbf{A}^{-1} + \mathbf{b}^T \mathbf{B}^{-1}) \underbrace{(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}}_{\mathbf{C}} \underbrace{(\mathbf{A}^{-1} + \mathbf{B}^{-1})}_{\mathbf{C}^{-1}} \mathbf{x} \\
&= (\mathbf{a}^T \mathbf{A}^{-1} + \mathbf{b}^T \mathbf{B}^{-1}) \underbrace{(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}}_{\mathbf{C}^T} \underbrace{(\mathbf{A}^{-1} + \mathbf{B}^{-1})}_{\mathbf{C}^{-1}} \mathbf{x} \\
&= \mathbf{c}^T \mathbf{C}^{-1} \mathbf{x}
\end{aligned}$$

Returning to (U) and making the replacements:

$$\begin{aligned}
&= -\frac{1}{2} \left[\mathbf{x}^T (\mathbf{A}^{-1} + \mathbf{B}^{-1}) \mathbf{x} - \mathbf{x}^T (\mathbf{A}^{-1} \mathbf{a} + \mathbf{B}^{-1} \mathbf{b}) - \right. \\
&\quad \left. (\mathbf{a}^T \mathbf{A}^{-1} + \mathbf{b}^T \mathbf{B}^{-1}) \mathbf{x} + \mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \mathbf{b}^T \mathbf{B}^{-1} \mathbf{b} \right] \tag{U} \\
&= -\frac{1}{2} \left[\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} - \mathbf{x}^T \mathbf{C}^{-1} \mathbf{c} - \mathbf{c}^T \mathbf{C}^{-1} \mathbf{x} + \mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \mathbf{b}^T \mathbf{B}^{-1} \mathbf{b} \right]
\end{aligned}$$

Completing the square:

$$\begin{aligned}
&= -\frac{1}{2} \left[\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} - \mathbf{x}^T \mathbf{C}^{-1} \mathbf{c} - \mathbf{c}^T \mathbf{C}^{-1} \mathbf{x} + \mathbf{c}^T \mathbf{C}^{-1} \mathbf{c} + \mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \mathbf{b}^T \mathbf{B}^{-1} \mathbf{b} - \mathbf{c}^T \mathbf{C}^{-1} \mathbf{c} \right] \\
&= -\frac{1}{2} \left[\mathbf{x}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{c}) - \mathbf{c}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{c}) + \mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \mathbf{b}^T \mathbf{B}^{-1} \mathbf{b} - \mathbf{c}^T \mathbf{C}^{-1} \mathbf{c} \right] \\
&= -\frac{1}{2} \left[(\mathbf{x} - \mathbf{c})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{c}) + \mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \mathbf{b}^T \mathbf{B}^{-1} \mathbf{b} - \mathbf{c}^T \mathbf{C}^{-1} \mathbf{c} \right] \\
&= -\frac{1}{2} \left[(\mathbf{x} - \mathbf{c})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{c}) \right] - \frac{1}{2} \left[\mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \mathbf{b}^T \mathbf{B}^{-1} \mathbf{b} - \mathbf{c}^T \mathbf{C}^{-1} \mathbf{c} \right]
\end{aligned}$$

The first term is exactly what we need for the exponent in $\mathcal{N}(\mathbf{c}, \mathbf{C})$. In order to get the required exponent in \mathbf{c} , we have to rewrite the second exponent a bit further.

We will need the following two results. The first one is a standard property of matrices:

$$\mathbf{X}^{-1} \mathbf{Y}^{-1} \mathbf{Z}^{-1} = [\mathbf{Z} \mathbf{Y} \mathbf{X}]^{-1}$$

The second one is a very unobvious equality:

$$\begin{aligned}
(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} &= \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1} \mathbf{A} \\
(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} &= \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1} \mathbf{B}
\end{aligned}$$

Will not give a formal proof of this, but some justification is needed.

By using some matrix algebra:

$$\begin{aligned}
(\mathbf{A} + \mathbf{B})^{-1} &= (\mathbf{B} + \mathbf{A})^{-1} \\
&= [(I + \mathbf{A}\mathbf{B}^{-1})\mathbf{B}]^{-1} \\
&= [\mathbf{A}(\mathbf{A}^{-1} + \mathbf{B}^{-1})\mathbf{B}]^{-1} \\
&= \mathbf{B}^{-1}(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}
\end{aligned}$$

The last equality follows from the result mentioned above. Now, we can solve the rest as a matrix equation and derive the equality:

$$\begin{aligned}
(\mathbf{A} + \mathbf{B})^{-1} &= \mathbf{B}^{-1}(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{A}^{-1} \\
\mathbf{B}(\mathbf{A} + \mathbf{B})^{-1} &= (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{A}^{-1} \\
\mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A} &= (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}
\end{aligned}$$

The symmetrical case is shown by not switching to $(\mathbf{B} + \mathbf{A})$ in the first line. Now, we can continue with the remainder from the exponent, and we replace the \mathbf{c} , \mathbf{c}^T and \mathbf{C} with the original values.

$$\begin{aligned}
&-\frac{1}{2}[\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} - \mathbf{c}^T\mathbf{C}^{-1}\mathbf{c}] \\
&= -\frac{1}{2}[\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} - (\mathbf{a}^T\mathbf{A}^{-1} + \mathbf{b}^T\mathbf{B}^{-1})\underbrace{(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}(\mathbf{A}^{-1} + \mathbf{B}^{-1})}_{=I}(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}(\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b})] \\
&= -\frac{1}{2}[\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} - (\mathbf{a}^T\mathbf{A}^{-1} + \mathbf{b}^T\mathbf{B}^{-1})(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}(\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b})]
\end{aligned}$$

To simplify notation, set $\mathbf{W} = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}$

$$\begin{aligned}
&= -\frac{1}{2}[\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} - (\mathbf{a}^T\mathbf{A}^{-1} + \mathbf{b}^T\mathbf{B}^{-1})\mathbf{W}(\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b})] \\
&= -\frac{1}{2}[\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} - \mathbf{a}^T\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{a} - \mathbf{a}^T\mathbf{A}^{-1}\mathbf{W}\mathbf{B}^{-1}\mathbf{b} - \mathbf{b}^T\mathbf{B}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{a} - \mathbf{b}^T\mathbf{B}^{-1}\mathbf{W}\mathbf{B}^{-1}\mathbf{b}] \\
&= -\frac{1}{2}[\mathbf{a}^T(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1})\mathbf{a} + \mathbf{b}^T(\mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{W}\mathbf{B}^{-1})\mathbf{b} - \mathbf{a}^T\mathbf{A}^{-1}\mathbf{W}\mathbf{B}^{-1}\mathbf{b} - \mathbf{b}^T\mathbf{B}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{a}] \quad (\#)
\end{aligned}$$

Now we can do some intermediary calculations. Treating each of these four terms separately. Note that we use $(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} = \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A}$.

$$\begin{aligned}
(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}) &= (\mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}) \\
&= \mathbf{A}^{-1}(I - (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}) \\
&= \mathbf{A}^{-1}(I - \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A}\mathbf{A}^{-1}) \\
&= \mathbf{A}^{-1}(I - \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}) \\
&= \mathbf{A}^{-1}((\mathbf{A} + \mathbf{B})(\mathbf{A} + \mathbf{B})^{-1} - \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}) \\
&= \mathbf{A}^{-1}(\mathbf{A} + \cancel{\mathbf{B}})(\mathbf{A} + \mathbf{B})^{-1} \\
&= \mathbf{A}^{-1}\mathbf{A}(\mathbf{A} + \mathbf{B})^{-1} \\
&= I(\mathbf{A} + \mathbf{B})^{-1} \\
&= (\mathbf{A} + \mathbf{B})^{-1}
\end{aligned}$$

Now we use the symmetrical result $(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} = \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}$.

$$\begin{aligned}
(\mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{W}\mathbf{B}^{-1}) &= (\mathbf{B}^{-1} - \mathbf{B}^{-1}(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{B}^{-1}) \\
&= \mathbf{B}^{-1}(I - (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{B}^{-1}) \\
&= \mathbf{B}^{-1}(I - \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B}\mathbf{B}^{-1}) \\
&= \mathbf{B}^{-1}(I - \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}) \\
&= \mathbf{B}^{-1}((\mathbf{A} + \mathbf{B})(\mathbf{A} + \mathbf{B})^{-1} - \mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}) \\
&= \mathbf{B}^{-1}(\mathbf{B} + \cancel{\mathbf{A}})(\mathbf{A} + \mathbf{B})^{-1} \\
&= \mathbf{B}^{-1}\mathbf{B}(\mathbf{A} + \mathbf{B})^{-1} \\
&= I(\mathbf{A} + \mathbf{B})^{-1} \\
&= (\mathbf{A} + \mathbf{B})^{-1}
\end{aligned}$$

The next cases are a little simpler.

$$\begin{aligned}
\mathbf{A}^{-1}\mathbf{W}\mathbf{B}^{-1} &= \mathbf{A}^{-1}(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{B}^{-1} \\
&= [\mathbf{B}(\mathbf{A}^{-1} + \mathbf{B}^{-1})\mathbf{A}]^{-1} \\
&= [\mathbf{B}(I + \mathbf{B}^{-1}\mathbf{A})]^{-1} \\
&= (\mathbf{B} + \mathbf{A})^{-1} \\
&= (\mathbf{A} + \mathbf{B})^{-1}
\end{aligned}$$

$$\begin{aligned}
\mathbf{B}^{-1}\mathbf{W}\mathbf{A}^{-1} &= \mathbf{B}^{-1}(\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{A}^{-1} \\
&= [\mathbf{A}(\mathbf{A}^{-1} + \mathbf{B}^{-1})\mathbf{B}]^{-1} \\
&= [\mathbf{A}(\mathbf{A}^{-1}\mathbf{B} + I)]^{-1} \\
&= (\mathbf{B} + \mathbf{A})^{-1} \\
&= (\mathbf{A} + \mathbf{B})^{-1}
\end{aligned}$$

All of these terms simplify to $(\mathbf{A} + \mathbf{B})^{-1}$! Returning to the main calculation.

$$\begin{aligned}
&= -\frac{1}{2}[\mathbf{a}^T(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1})\mathbf{a} + \mathbf{b}^T(\mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{W}\mathbf{B}^{-1})\mathbf{b} - \mathbf{a}^T\mathbf{A}^{-1}\mathbf{W}\mathbf{B}^{-1}\mathbf{b} - \mathbf{b}^T\mathbf{B}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{a}] \quad (\#) \\
&= -\frac{1}{2}[\mathbf{a}^T(\mathbf{A} + \mathbf{B})^{-1}\mathbf{a} + \mathbf{b}^T(\mathbf{A} + \mathbf{B})^{-1}\mathbf{b} - \mathbf{a}^T(\mathbf{A} + \mathbf{B})^{-1}\mathbf{b} - \mathbf{b}^T(\mathbf{A} + \mathbf{B})^{-1}\mathbf{a}] \\
&= -\frac{1}{2}[\mathbf{a}^T(\mathbf{A} + \mathbf{B})^{-1}(\mathbf{a} - \mathbf{b}) + \mathbf{b}^T(\mathbf{A} + \mathbf{B})^{-1}\mathbf{b} - \mathbf{b}^T(\mathbf{A} + \mathbf{B})^{-1}\mathbf{a}] \\
&= -\frac{1}{2}[\mathbf{a}^T(\mathbf{A} + \mathbf{B})^{-1}(\mathbf{a} - \mathbf{b}) - \mathbf{b}^T(\mathbf{A} + \mathbf{B})^{-1}(\mathbf{a} - \mathbf{b})] \\
&= -\frac{1}{2}(\mathbf{a} - \mathbf{b})^T(\mathbf{A} + \mathbf{B})^{-1}(\mathbf{a} - \mathbf{b})
\end{aligned}$$

With this we have finally resolved the exponent of the product.

Now that we have completed the exponents $W_1 + W_2$, we turn our attention to the normalizing factors $N_1 N_2$. We will use the following two properties:

$$|A||B| = |AB|, \quad \text{and} \quad |A^{-1}| = \frac{1}{|A|}$$

From the simplified equation we defined earlier:

$$\begin{aligned} N_1 N_2 &= \left((2\pi)^{-\frac{D}{2}} |\mathbf{A}|^{-\frac{1}{2}} \right) \left((2\pi)^{-\frac{D}{2}} |\mathbf{B}|^{-\frac{1}{2}} \right) \\ &= (2\pi)^{-\frac{D}{2}} (2\pi)^{-\frac{D}{2}} |\mathbf{A}|^{-\frac{1}{2}} |\mathbf{B}|^{-\frac{1}{2}} \\ &= (2\pi)^{-\frac{D}{2}} (2\pi)^{-\frac{D}{2}} |\mathbf{C}|^{-\frac{1}{2}} \frac{|\mathbf{A}|^{-\frac{1}{2}} |\mathbf{B}|^{-\frac{1}{2}}}{|\mathbf{C}|^{-\frac{1}{2}}} \\ &= (2\pi)^{-\frac{D}{2}} |\mathbf{C}|^{-\frac{1}{2}} (2\pi)^{-\frac{D}{2}} |\mathbf{A}|^{-\frac{1}{2}} |\mathbf{C}^{-1}|^{-\frac{1}{2}} |\mathbf{B}|^{-\frac{1}{2}} \\ &= (2\pi)^{-\frac{D}{2}} |\mathbf{C}|^{-\frac{1}{2}} (2\pi)^{-\frac{D}{2}} |\mathbf{A}\mathbf{C}^{-1}\mathbf{B}|^{-\frac{1}{2}} \\ &= (2\pi)^{-\frac{D}{2}} |\mathbf{C}|^{-\frac{1}{2}} (2\pi)^{-\frac{D}{2}} |\mathbf{A}(\mathbf{A}^{-1} + \mathbf{B}^{-1})\mathbf{B}|^{-\frac{1}{2}} \\ &= (2\pi)^{-\frac{D}{2}} |\mathbf{C}|^{-\frac{1}{2}} (2\pi)^{-\frac{D}{2}} |\mathbf{A} + \mathbf{B}|^{-\frac{1}{2}} \end{aligned}$$

With this we have separated $N_1 N_2$ into the two desired parts. In conclusion, we have showed that:

$$\begin{aligned} p(\mathbf{x}|\mathbf{a}, \mathbf{A})p(\mathbf{x}|\mathbf{b}, \mathbf{B}) &= N_1 \exp(W_1)N_2 \exp(W_2) = N_1 N_2 \exp(W_1 + W_2) \\ &= c(2\pi)^{-\frac{D}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mathbf{c})^T \mathbf{C}^{-1}(\mathbf{x} - \mathbf{c}) \right\} \end{aligned}$$

and where c is:

$$c = (2\pi)^{-\frac{D}{2}} |\mathbf{A} + \mathbf{B}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(\mathbf{a} - \mathbf{b})^T (\mathbf{A} + \mathbf{B})^{-1}(\mathbf{a} - \mathbf{b}) \right\}$$

b.

For sake of completeness, we will first consider the exponential form of a univariate Gaussian distribution. There we have:

$$h(\mathbf{x}) = \frac{1}{\sqrt{2\pi}}, \quad \phi(\mathbf{x}) = \begin{bmatrix} x \\ x^2 \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \mu/\sigma^2 \\ -1/\sigma^2 \end{bmatrix}, \quad A(\boldsymbol{\theta}) = \frac{\mu^2}{2\sigma^2} + \log \sigma$$

Will need that $e^{-\log \sigma} = 1/e^{\log \sigma} = 1/\sigma$. Then:

$$\begin{aligned} h(\mathbf{x}) \exp \left[\phi(\mathbf{x})^T \boldsymbol{\theta} - A(\boldsymbol{\theta}) \right] &= \frac{1}{\sqrt{2\pi}} \exp \left\{ \frac{\mu x}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - \log \sigma \right\} \\ &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (x^2 - 2\mu x + \mu^2) \right\} \exp \{-\log \sigma\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} \sim \mathcal{N}(\mu, \sigma^2) \end{aligned}$$

Here is the general exponential form for a multivariate Gaussian distribution. Will use without deriving it - that's too much work! Note the similarities to the univariate exponential form.

$$h(\mathbf{x}) = (2\pi)^{-\frac{D}{2}}, \quad \phi(\mathbf{x}) = \begin{bmatrix} \mathbf{x} \\ \text{vec}(\mathbf{x}\mathbf{x}^T) \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ \text{vec}(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}) \end{bmatrix}, \quad A(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{1}{2}\log|\boldsymbol{\Sigma}|$$

Here $\text{vec}(\cdot)$ turns an $(m \times n)$ matrix into a $(mn \times 1)$ vector by stacking the columns. We have verified this in the 2D multivariate case in the **Appendix**. Writing the two distributions on the exponential form (after moving the covariance matrix out of the exponent):

$$p(\mathbf{x}|\mathbf{a}, \mathbf{A}) = (2\pi)^{-\frac{D}{2}} |\mathbf{A}|^{-\frac{1}{2}} \exp \left\{ \begin{bmatrix} \mathbf{x} \\ \text{vec}(\mathbf{x}\mathbf{x}^T) \end{bmatrix}^T \begin{bmatrix} \mathbf{A}^{-1}\mathbf{a} \\ \text{vec}(-\frac{1}{2}\mathbf{A}^{-1}) \end{bmatrix} - \frac{1}{2}\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} \right\}$$

$$p(\mathbf{x}|\mathbf{b}, \mathbf{B}) = (2\pi)^{-\frac{D}{2}} |\mathbf{B}|^{-\frac{1}{2}} \exp \left\{ \begin{bmatrix} \mathbf{x} \\ \text{vec}(\mathbf{x}\mathbf{x}^T) \end{bmatrix}^T \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \text{vec}(-\frac{1}{2}\mathbf{B}^{-1}) \end{bmatrix} - \frac{1}{2}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} \right\}$$

If we consider the shorthand version of this as we did in (a), we get:

$$p(\mathbf{x}|\mathbf{a}, \mathbf{A})p(\mathbf{x}|\mathbf{b}, \mathbf{B}) = N_1 \exp(W_1)N_2 \exp(W_2) = N_1N_2 \exp(W_1 + W_2)$$

Note first that the normalizing factor N_1N_2 follows the same argument as on the previous page, so all we need to do is focus on the sum of the exponents. From the inner products we get:

$$\begin{bmatrix} \mathbf{x} \\ \text{vec}(\mathbf{x}\mathbf{x}^T) \end{bmatrix}^T \begin{bmatrix} \mathbf{A}^{-1}\mathbf{a} \\ \text{vec}(-\frac{1}{2}\mathbf{A}^{-1}) \end{bmatrix} = \mathbf{x}^T\mathbf{A}^{-1}\mathbf{a} + \text{vec}(\mathbf{x}\mathbf{x}^T)^T \text{vec}(-\frac{1}{2}\mathbf{A}^{-1})$$

$$\begin{bmatrix} \mathbf{x} \\ \text{vec}(\mathbf{x}\mathbf{x}^T) \end{bmatrix}^T \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \text{vec}(-\frac{1}{2}\mathbf{B}^{-1}) \end{bmatrix} = \mathbf{x}^T\mathbf{B}^{-1}\mathbf{b} + \text{vec}(\mathbf{x}\mathbf{x}^T)^T \text{vec}(-\frac{1}{2}\mathbf{B}^{-1})$$

So if we ignore the exponent and only focus on $W_1 + W_2$, we have the following sum:

$$\mathbf{x}^T\mathbf{A}^{-1}\mathbf{a} + \text{vec}(\mathbf{x}\mathbf{x}^T)^T \text{vec}(-\frac{1}{2}\mathbf{A}^{-1}) - \frac{1}{2}\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} + \mathbf{x}^T\mathbf{B}^{-1}\mathbf{b} + \text{vec}(\mathbf{x}\mathbf{x}^T)^T \text{vec}(-\frac{1}{2}\mathbf{B}^{-1}) - \frac{1}{2}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{b}$$

As seen in the simplified example in the **Appendix** (but not showing formally), we can write:

$$\text{vec}(\mathbf{x}\mathbf{x}^T)^T \text{vec}(-\frac{1}{2}\mathbf{A}^{-1}) = -\frac{1}{2}\mathbf{x}^T\mathbf{A}^{-1}\mathbf{x}$$

and,

$$\text{vec}(\mathbf{x}\mathbf{x}^T)^T \text{vec}(-\frac{1}{2}\mathbf{B}^{-1}) = -\frac{1}{2}\mathbf{x}^T\mathbf{B}^{-1}\mathbf{x}.$$

For all parts containing \mathbf{A} we can use the other rewriting trick we used in the simplified case:

$$\begin{aligned} \mathbf{x}^T\mathbf{A}^{-1}\mathbf{a} - \frac{1}{2}\mathbf{x}^T\mathbf{A}^{-1}\mathbf{x} - \frac{1}{2}\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} &= -\frac{1}{2}[\mathbf{x}^T\mathbf{A}^{-1}\mathbf{x} - 2\mathbf{x}^T\mathbf{A}^{-1}\mathbf{a} - \frac{1}{2}\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a}] \\ &= -\frac{1}{2}(\mathbf{x} - \mathbf{a})^T\mathbf{A}^{-1}(\mathbf{x} - \mathbf{a}) \end{aligned}$$

The case for the terms with \mathbf{B} is exactly the same.

$$\begin{aligned} \mathbf{x}^T\mathbf{B}^{-1}\mathbf{b} - \frac{1}{2}\mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} - \frac{1}{2}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} &= -\frac{1}{2}[\mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} - 2\mathbf{x}^T\mathbf{B}^{-1}\mathbf{b} - \frac{1}{2}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{b}] \\ &= -\frac{1}{2}(\mathbf{x} - \mathbf{b})^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{b}) \end{aligned}$$

With these, we have rewritten the exponential forms back to the original multivariate Gaussian exponents, and following the calculations in part (a) we can split up the product in the intended way and get $c\mathcal{N}(\mathbf{x}|\mathbf{c}, \mathbf{C})$ in the same way.

I'm wondering if there is an easier way of doing this, since calculating the product from the exponential form just introduces some extra steps. But I cannot see an easier way.

6.11. Iterated expectation. Showing this for the continuous case. For the discrete we can just replace the integrals with sums. Defining \mathbf{X} and \mathbf{Y} as all possible values for x and y (such as \mathbb{R}). Assuming that the expectations exist which means we can change the order of the integrals (or sums). We will use the product rule (P.R):

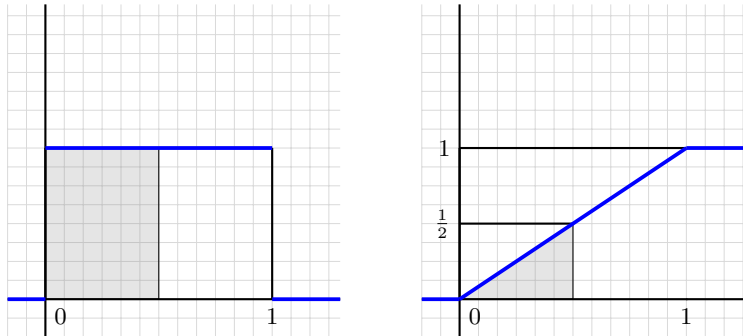
$$p(x, y) = p(x|y)p(y)$$

and the sum rule (S.R):

$$\begin{aligned} p(x) &= \int_{\mathbf{Y}} p(x, y) dy \\ \mathbb{E}_{\mathbf{Y}}[\mathbb{E}_{\mathbf{X}}[x|y]] &= \mathbb{E}_{\mathbf{Y}} \left[\int_{\mathbf{X}} x \cdot p(x|y) dx \right] \\ &= \int_{\mathbf{Y}} \int_{\mathbf{X}} x \cdot p(x|y) dx p(y) dy \\ &= \int_{\mathbf{Y}} \int_{\mathbf{X}} x \cdot p(x|y)p(y) dx dy \\ &\stackrel{P.R}{=} \int_{\mathbf{Y}} \int_{\mathbf{X}} x \cdot p(x, y) dx dy \\ &= \int_{\mathbf{X}} x \int_{\mathbf{Y}} p(x, y) dy dx \\ &\stackrel{S.R}{=} \int_{\mathbf{X}} x \cdot p(x) dx \\ &= \mathbb{E}_{\mathbf{X}}[x] \end{aligned}$$

6.12 Will solve later.

6.13. This is an important result for Monte Carlo simulations that allow us to turn uniformly distributed values from $u \in [0, 1]$ into variables from many distributions by calculating $F^{-1}(u)$. First, a short reminder of the pdf and cdf for a uniform distribution on $[0, 1]$.



If we have some $Z \sim U(0, 1)$, then $P(Z \leq 0.5)$ is the area marked in gray on the left image above. The probability will be $P(Z \leq 0.5) = 0.5$. The same is true for any value in $[0, 1]$. This is an important property for $U(0, 1)$ and is what we are going to prove.

Theorem 6.15

Let X be a continuous random variable with a strictly monotonic cumulative distribution function $F_X(x)$. Then the random variable Y defined as

$$Y := F_X(X) \tag{6.132}$$

has a uniform distribution.

Proof.

Assume $y \in (0, 1)$. We want to show that $P(Y \leq y) = y$. Since $F_X(x)$ is strictly monotonic on $(0, 1)$, it is injective, so the inverse function is well defined.

$$\begin{aligned} P(Y \leq y) &= P(F_X(X) \leq y) \\ &= P(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) \\ &= y \end{aligned}$$

So for $y \in (0, 1)$, we have shown that $P(Y \leq y) = y$. For $y \leq 0$ then $P(Y \leq y) = 0$ and for $y \geq 1$ then $P(Y \leq y) = 1$. This proves that Y has a uniform distribution. ■

Chapter 7: Continuous Optimization - Exercises

7.1. Finding and classifying the stationary points of:

$$f(x) = x^3 + 6x^2 - 3x - 5.$$

First, we calculate the first and second derivative:

$$\begin{aligned}f'(x) &= 3x^2 + 12x - 3 \\f''(x) &= 6x + 12\end{aligned}$$

The stationary points are where the derivative is 0.

$$\begin{aligned}f'(x) &= 0 \\3x^2 + 12x - 3 &= 0 \\x^2 + 4x - 1 &= 0 \\(x + 2 - \sqrt{5})(x + 2 + \sqrt{5}) &= 0\end{aligned}$$

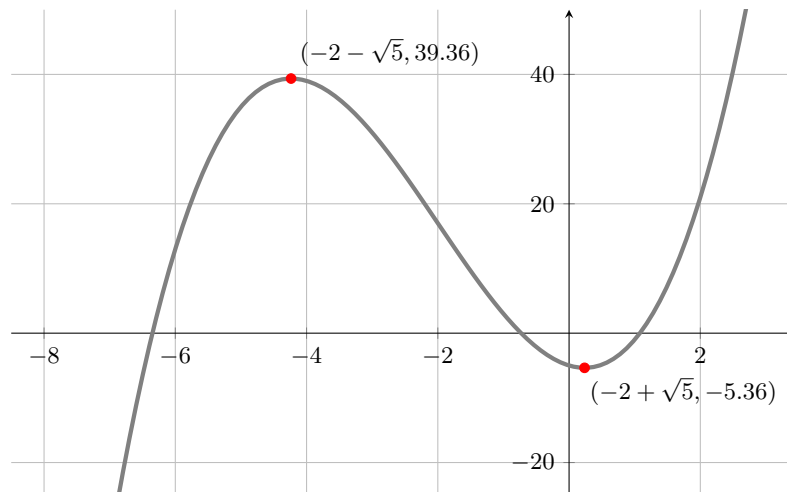
So there are stationary points at: $x_1 = -2 + \sqrt{5}$ and $x_2 = -2 - \sqrt{5}$. Inserting these points into $f''(x)$ in order to classify the points.

$$\begin{aligned}f''(x_1) &= 6(-2 + \sqrt{5}) + 12 = -12 + 6\sqrt{5} + 12 = 6\sqrt{5} > 0 \\f''(x_2) &= 6(-2 - \sqrt{5}) + 12 = -12 - 6\sqrt{5} + 12 = -6\sqrt{5} < 0\end{aligned}$$

Positive second derivative means x_1 is a local minimum: $f(x_1) \approx -5.36$.

Negative second derivative means x_2 is a local maximum: $f(x_2) \approx 39.36$

Illustration of the function $f(x) = x^3 + 6x^2 - 3x - 5$:



7.2

In gradient descent we want to minimize some objective function $L(\cdot)$ by using the parameter θ (which could be the parameters of some model, like α and β in a simple linear regression).

In gradient descent we start at some random point and update the parameters according to the gradient of L which decides the DIRECTION we will go, and according to some step-size parameter γ , which decides how FAR we will go.

In standard gradient descent, this is done for all $1, \dots, N$ data points:

$$\theta_{i+1} = \theta_i - \gamma_i (\nabla L(\theta_i))^T = \theta_i - \gamma_i \sum_{n=1}^N (\nabla L_n(\theta_i))^T \quad (7.15)$$

As discussed in the text book, this can often be simplified by using mini-batches which uses $k < N$ points instead of doing the full calculations. Very often this will lead to the same solution with the advantage of being a lot computationally cheaper to do. In stochastic gradient descent, with $k = 1$ we could rewrite the formula as follows:

$$\text{For some } \lambda \in \{1, \dots, N\}, \quad \theta_{i+1} = \theta_i - \gamma_i (\nabla L_\lambda(\theta_i))^T$$

We update the parameters wrt. a single L_λ .

7.3. Considering the validity of statements on convex sets.

a. **Statement:** The intersection of any two convex sets is convex.

This is true, which we can prove.

Proof.

Assume A and B are convex sets. Now assume $x, y \in A \cap B$. Since A is convex then for any $\theta \in [0, 1]$ we have $\theta x + (1 - \theta)y \in A \cap B$. The same property follows from B . Hence, $A \cap B$ is a convex set. ■

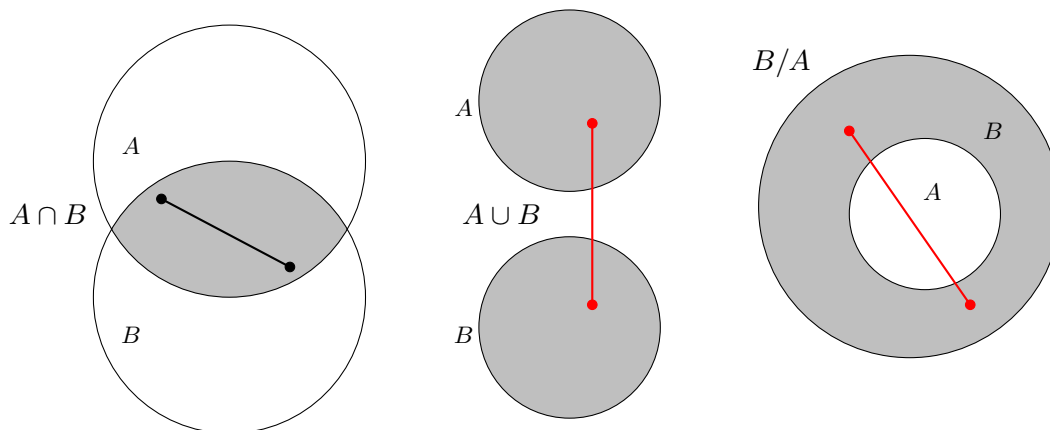
b. **Statement:** The union of any two convex sets is convex.

This is false. Counterexample: disjoint circles.

c. **Statement:** The difference of a convex set A from another convex set B is convex.

This is false. Counterexample: A is a subset of B .

Illustrations. Red lines show examples of where the convexity property is violated, since the convexity property requires the line between two points in the set to also be contained in the set.



7.4. Considering the validity of statements on convex functions.

a. Statement: The sum of any two convex functions is convex.

This is true. Will not prove this here since it was already proved in the text book in example 7.4. (Equations (7.34) to (7.37)).

b. Statement: The difference of any two convex functions is convex.

This is false. Consider the two convex functions $f(x) = x^2$ and $g(x) = 2x^2$, then $h(x) = f(x) - g(x) = -x^2$ which is equal to $-f(x)$ and therefore a concave function.

c. Statement: The product of any two convex functions is convex.

This is false. Set $f(x) = x^2 + 2x$ and $g(x) = x^2 - 2x$ which are both convex. Then the product is:

$$h(x) = f(x)g(x) = x^4 - 4x^2 = x^2(x^2 - 4).$$

By setting $\theta = \frac{1}{2}$ and $x = -1$ and $y = 1$ we get:

$$\begin{aligned} h(\theta x + (1 - \theta)y) &= h\left(\frac{1}{2}(-1) + \frac{1}{2}(1)\right) = h(0) = 0 \\ \theta h(x) + (1 - \theta)h(y) &= \frac{1}{2}h(1) + \frac{1}{2}h(-1) = \frac{1}{2}(-3) + \frac{1}{2}(-3) = -3 \end{aligned}$$

So, $\theta h(x) + (1 - \theta)h(y) < h(\theta x + (1 - \theta)y)$ which shows that the product $h(x)$ is not convex.

d. Statement: The maximum of any two convex functions is convex.

This is true.

Proof.

Assume f and g are convex functions and define $h(\mathbf{x}) := \max(f(\mathbf{x}), g(\mathbf{x}))$. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^D$. Also assume, without loss of generality that

$$g(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) \leq f(\theta \mathbf{x} + (1 - \theta)\mathbf{y}).$$

(Just name the functions accordingly). Then:

$$\begin{aligned} h(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) &= \max(f(\theta \mathbf{x} + (1 - \theta)\mathbf{y}), g(\theta \mathbf{x} + (1 - \theta)\mathbf{y})) \\ &= f(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) \\ &\leq \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}) \\ &\leq \theta \max(f(\mathbf{x}), g(\mathbf{x})) + (1 - \theta) \max(f(\mathbf{y}), g(\mathbf{y})) \\ &= \theta h(\mathbf{x}) + (1 - \theta)h(\mathbf{y}) \end{aligned}$$

In conclusion, this shows that h satisfies the condition for convexity:

$$h(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) \leq \theta h(\mathbf{x}) + (1 - \theta)h(\mathbf{y}) \quad \blacksquare$$

7.5. Rewriting the optimization problem in standard matrix form. (We will use x_1, x_2 instead of x_0, x_1 .) Start by noting that:

$$\mathbf{p}^T \mathbf{x} + \xi = p_1 x_1 + p_2 x_2 + (1)\xi,$$

so we can set $\mathbf{q} \in \mathbb{R}^3$ as $[p_1 \ p_2 \ 1]$ and set $\mathbf{x} \in \mathbb{R}^3$ as $[x_1 \ x_2 \ \xi]$.

Next are the constraints. $x_1 \leq 0$, $x_2 \leq 3$ and $\xi \geq 0$, or equivalently, $-\xi \leq 0$, which we can easily write in matrix form. We get:

$$\begin{aligned} & \max_{\mathbf{x} \in \mathbb{R}^3, \xi \in \mathbb{R}} \begin{bmatrix} p_1 \\ p_2 \\ 1 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ \xi \end{bmatrix} \\ & \text{subject to } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \xi \end{bmatrix} \leq \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} \end{aligned}$$

If we make the following definitions:

$$\mathbf{c} = \begin{bmatrix} p_1 \\ p_2 \\ 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \xi \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix},$$

we can formulate the original problem in the standard, matrix form:

$$\begin{aligned} & \max_{\mathbf{x} \in \mathbb{R}^3, \xi \in \mathbb{R}} \mathbf{c}^T \mathbf{x} \\ & \text{subject to } \mathbf{A} \mathbf{x} \leq \mathbf{b} \end{aligned}$$

7.6. Deriving the dual linear program by using the Lagrange duality. We are given the following primal problem:

$$\begin{aligned} & \min_{\mathbf{x} \in \mathbb{R}^2} - \begin{bmatrix} 5 \\ 3 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ & \text{subject to } \begin{bmatrix} 2 & 2 \\ 2 & -4 \\ -2 & 1 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix} \end{aligned}$$

When we have defined an optimization problem in the standard matrix notation, we can easily convert it to the dual problem like this:

PRIMAL PROBLEM

$$\begin{aligned} & \min_{\mathbf{x} \in \mathbb{R}^D} \mathbf{c}^T \mathbf{x} \\ & \text{subject to } \mathbf{A} \mathbf{x} \leq \mathbf{b} \end{aligned}$$

DUAL PROBLEM

$$\begin{aligned} & \implies \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} -\mathbf{b}^T \boldsymbol{\lambda} \\ & \text{subject to } \mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda} = \mathbf{0} \\ & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned}$$

By just filling in the vectors, we get the dual problem.

$$\max_{\lambda \in \mathbb{R}^5} - \begin{bmatrix} 33 \\ 8 \\ 5 \\ -1 \\ 8 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix}$$

$$\text{subject to } - \begin{bmatrix} 5 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 & 2 & -2 & 0 & 0 \\ 2 & -4 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$\lambda \geq \mathbf{0}$$

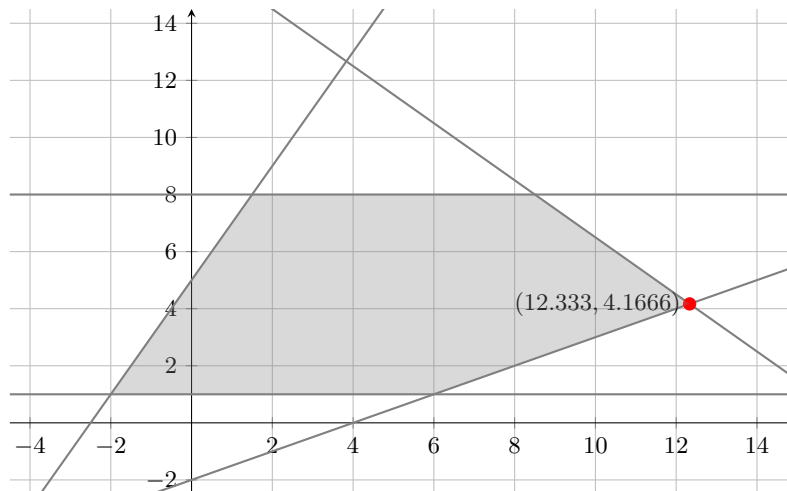
Both these problems were solved in Python (see **Appendix**), and we verified that the primal problem has the minimal value -74.167 for values $x_1 = \frac{37}{3} \approx 12.333$ and $x_2 = \frac{25}{6} \approx 4.167$ (as shown in the plot). Verifying:

$$\mathbf{c}^T \mathbf{x} = (-5) \left(\frac{37}{3} \right) + (-3) \left(\frac{25}{6} \right) = -\frac{445}{6} = -74.1667$$

The dual problem has the same solution with $\lambda_1 = \frac{13}{6}$ and $\lambda_2 = \frac{1}{3}$ and $\lambda_3 = \lambda_4 = \lambda_5 = 0$, which we can verify gives the same value:

$$-\mathbf{b}^T \boldsymbol{\lambda} = (-33) \left(\frac{13}{6} \right) + (-8) \left(\frac{1}{3} \right) = -\frac{445}{6} = -74.1667$$

Illustration of the primal problem, with the optimal point marked in red. (Not possible to visualize the dual problem as it is in 5D-space).



7.7. We have the following quadratic program:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^2} \quad & \frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 5 \\ 3 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ \text{subject to} \quad & \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \end{aligned}$$

From the text book, we know how to go from the primal problem to the dual problem, as outlined in equation (7.45) and (7.52).

PRIMAL PROBLEM

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^D} \quad & \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x} \\ \text{subject to} \quad & \mathbf{A} \mathbf{x} \leq \mathbf{b} \end{aligned}$$

DUAL PROBLEM

$$\begin{aligned} \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \quad & -\frac{1}{2} (\mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda})^T \mathbf{Q}^{-1} (\mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda}) - \mathbf{b}^T \boldsymbol{\lambda} \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq 0 \end{aligned}$$

In our case, $\boldsymbol{\lambda} \in \mathbb{R}^4$. Some calculations (skipping the details):

$$\begin{aligned} \mathbf{A}^T &= \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \\ \mathbf{Q}^{-1} &= \begin{bmatrix} 4/7 & -1/7 \\ -1/7 & 2/7 \end{bmatrix} \\ \mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda} &= \begin{bmatrix} 5 + \lambda_1 - \lambda_2 \\ 3 + \lambda_3 - \lambda_4 \end{bmatrix} \end{aligned}$$

Filling in all the values lets us express the dual quadratic problem.

$$\begin{aligned} \max_{\boldsymbol{\lambda} \in \mathbb{R}^4} \quad & -\frac{1}{2} \begin{bmatrix} 5 + \lambda_1 - \lambda_2 \\ 3 + \lambda_3 - \lambda_4 \end{bmatrix}^T \begin{bmatrix} 4/7 & -1/7 \\ -1/7 & 2/7 \end{bmatrix} \begin{bmatrix} 5 + \lambda_1 - \lambda_2 \\ 3 + \lambda_3 - \lambda_4 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq 0. \end{aligned}$$

7.8. We have another quadratic programming problem, and note that we are minimizing over \mathbf{w} :

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^D} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^T \mathbf{x} \geq 1 \end{aligned}$$

Written out in more detail:

$$\begin{aligned} \frac{1}{2} \mathbf{w}^T \mathbf{w} &= \frac{1}{2} (w_1^2 + \dots + w_D^2) \\ \mathbf{w}^T \mathbf{x} &= w_1 x_1 + \dots + w_D x_D \geq 1 \end{aligned}$$

We are minimizing over \mathbf{w} and we can rewrite the last expression as $\mathbf{x}^T \mathbf{w}$ since $\mathbf{x}^T \mathbf{w} = \mathbf{w}^T \mathbf{x}$.

Our reformulated problem.

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^D} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} \\ \text{subject to} \quad & -\mathbf{x}^T \mathbf{w} \leq -1 \end{aligned}$$

We can rewrite this to the standard primal problem matrix form by making the following observations, and changing the given constraint to $-\mathbf{x}^T \mathbf{w} \leq -1$:

$$\mathbf{Q} = I_D, \quad \mathbf{c} = \mathbf{0}, \quad \mathbf{A} = -\mathbf{x}^T \quad \mathbf{b} = -1$$

Since we have a one-dimensional constraint, we get $\boldsymbol{\lambda} = \lambda$, that is, λ is not a vector, but a number $\lambda \in \mathbb{R}$. Some intermediary calculations:

$$\begin{aligned} \boldsymbol{\lambda} &= \lambda \\ \mathbf{A}^T &= (-\mathbf{x}^T)^T = -\mathbf{x} \\ \mathbf{Q}^{-1} &= I_D \\ \mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda} &= \mathbf{0} - \mathbf{x} \lambda = -\mathbf{x} \lambda \end{aligned}$$

Using these values in the general dual problem:

$$\begin{aligned} \max_{\boldsymbol{\lambda} \in \mathbb{R}^m} \quad & -\frac{1}{2} (\mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda})^T \mathbf{Q}^{-1} (\mathbf{c} + \mathbf{A}^T \boldsymbol{\lambda}) - \mathbf{b}^T \boldsymbol{\lambda} \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned}$$

Which leaves us with:

$$\begin{aligned} \max_{\lambda \in \mathbb{R}} \quad & -\frac{1}{2} (-\mathbf{x} \lambda)^T (-\mathbf{x} \lambda) + \lambda \\ \text{subject to} \quad & \lambda \geq 0 \end{aligned}$$

Since λ is a real number, it is not affected by the transpose, and we can factor it out. The negatives also cancel each other. The dual problem can therefore be written as:

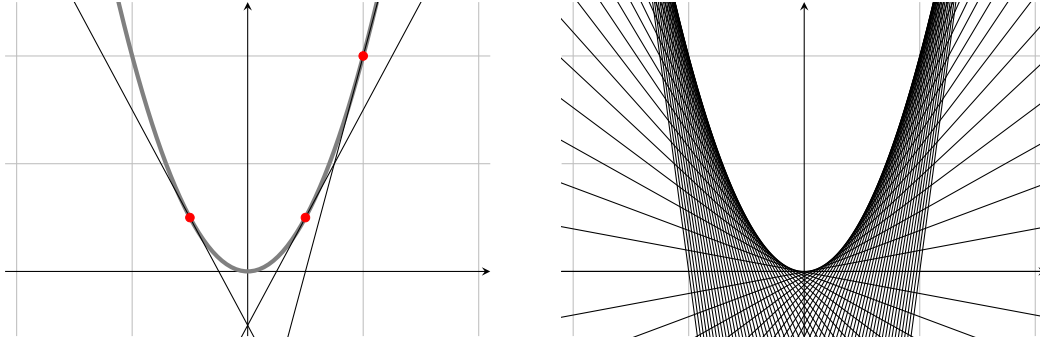
$$\begin{aligned} \max_{\lambda \in \mathbb{R}} \quad & -\frac{\lambda^2}{2} \mathbf{x}^T \mathbf{x} + \lambda \\ \text{subject to} \quad & \lambda \geq 0 \end{aligned}$$

7.9. Finding the convex conjugate function $f^*(s)$ of the negative entropy of $\mathbf{x} \in \mathbb{R}^D$:

$$f(\mathbf{x}) = \sum_{i=1}^D x_i \log x_i.$$

Assuming the standard dot product. Will illustrate the convex conjugate a bit further. See plots on next page.

Graphical illustration of the convex conjugate for $f(x) = x^2$. On the left, there are three specific tangents. On the right there are 80 tangent lines. In 2D they are lines, but more generally they are supporting hyperplanes, and can be used to describe the convex set created by the convex function $f(x) = x^2$, as we can see.



In the left image we have three tangent lines with slope s . The supporting hyperplanes are lines of the form $y = sx + c$ and the minimal point where these lines touches the convex set are marked with red points (which are simply the tangent points). The minimal value c where the lines touches the function graph is $y = f(x_0)$. By setting this equal to the value of the line, we get:

$$\left. \begin{array}{l} y = sx_0 + c \\ y = f(x_0) \end{array} \right\} \implies c = \inf_{x_0} [-sx_0 + f(x_0)]$$

which is a description of the line that touches the convex function graph. The 'opposite' of this is the definition of the convex conjugate (for 1 dimensions).

$$f^*(s) = \sup_x [sx - f(x)]$$

Now, to the exercise. By the definition of the convex conjugate:

$$\begin{aligned} f^*(\mathbf{s}) &= \sup_{\mathbf{x}} \left\{ \mathbf{s}^T \mathbf{x} - f(\mathbf{x}) \right\} \\ &= \sup_{\mathbf{x}} \left\{ \mathbf{s}^T \mathbf{x} - \sum_{i=1}^D x_i \log x_i \right\} \\ &= \sup_{\mathbf{x}} \left\{ \sum_{i=1}^D s_i x_i - x_i \log x_i \right\} \end{aligned}$$

Key observation: we can put the supremum inside the sum.

$$= \sum_{i=1}^D \left(\sup_{x_i} \{ s_i x_i - x_i \log x_i \} \right) \quad (\heartsuit)$$

We can investigate each term separately. In order to find the x_i that maximizes each term, we differentiate it and set it equal to 0.

$$\frac{d}{dx_i} (s_i x_i - x_i \log x_i) = s_i - ((1) \log x_i + x_i \cdot \frac{1}{x_i}) = s_i - \log x_i - 1$$

Setting equal to 0 and solving for x_i .

$$\begin{aligned} s_i - \log x_i - 1 &= 0 \\ \log x_i &= s_i - 1 \\ x_i &= \exp\{s_i - 1\} \end{aligned}$$

This x_i maximizes the expression in (\heartsuit). Replacing it back in:

$$\begin{aligned} f^*(\mathbf{s}) &= \sum_{i=1}^D \left(\sup_{x_i} \{s_i x_i - x_i \log x_i\} \right) \\ &= \sum_{i=1}^D (s_i \exp\{s_i - 1\} - \exp\{s_i - 1\} \log \exp\{s_i - 1\}) \\ &= \sum_{i=1}^D (s_i \exp\{s_i - 1\} - \exp\{s_i - 1\} (s_i - 1)) \\ &= \sum_{i=1}^D (\exp\{s_i - 1\} (s_i - (s_i - 1))) \\ &= \sum_{i=1}^D (\exp\{s_i - 1\} (s_i - s_i + 1)) \\ &= \sum_{i=1}^D \exp\{s_i - 1\} \end{aligned}$$

This function is the convex conjugate to the negative entropy.

7.10. Finding the convex conjugate of:

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c,$$

where \mathbf{A} is strictly positive definite, which means it is symmetric, so it is invertible and $A^T = A$ and $(A^{-1})^T = A^{-1}$. By the definition of the convex conjugate:

$$\begin{aligned} f^*(\mathbf{s}) &= \sup_{\mathbf{x}} \{ \mathbf{s}^T \mathbf{x} - f(\mathbf{x}) \} \\ &= \sup_{\mathbf{x}} \left\{ \mathbf{s}^T \mathbf{x} - \left(\frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c \right) \right\} \\ &= \sup_{\mathbf{x}} \left\{ \mathbf{s}^T \mathbf{x} - \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x} - c \right\} \end{aligned} \quad (\star)$$

To find the \mathbf{x} that maximizes this, we differentiate it and set it equal to 0. Without going into the details:

$$\frac{d}{d\mathbf{x}} \left(\mathbf{s}^T \mathbf{x} - \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x} - c \right) = \mathbf{s}^T - \mathbf{A} \mathbf{x} - \mathbf{b}^T = \mathbf{s} - \mathbf{A} \mathbf{x} - \mathbf{b}$$

Setting equal to 0 and solving for \mathbf{x} .

$$\begin{aligned}\mathbf{s} - \mathbf{Ax} - \mathbf{b} &= \mathbf{0} \\ \mathbf{Ax} &= \mathbf{s} - \mathbf{b} \\ \mathbf{x} &= \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b})\end{aligned}$$

Setting this in for \mathbf{x} in (★) (and removing the supremum). Recall that \mathbf{A} is symmetric, so $(\mathbf{A}^{-1})^T = \mathbf{A}^{-1}$.

$$\begin{aligned}f^*(\mathbf{s}) &= \sup_{\mathbf{x}} \left\{ \mathbf{s}^T \mathbf{x} - \frac{1}{2} \mathbf{x}^T \mathbf{Ax} - \mathbf{b}^T \mathbf{x} - c \right\} && (\star) \\ &= \mathbf{s}^T \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}) - \frac{1}{2} (\mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}))^T \mathbf{AA}^{-1}(\mathbf{s} - \mathbf{b}) - \mathbf{b}^T \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}) - c \\ &= \mathbf{s}^T \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}) - \frac{1}{2} (\mathbf{s} - \mathbf{b})^T \mathbf{A}^{-1} \mathbf{AA}^{-1}(\mathbf{s} - \mathbf{b}) - \mathbf{b}^T \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}) - c \\ &= \mathbf{s}^T \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}) - \mathbf{b}^T \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}) - \frac{1}{2} (\mathbf{s} - \mathbf{b})^T \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}) - c \\ &= (\mathbf{s} - \mathbf{b})^T \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}) - \frac{1}{2} (\mathbf{s} - \mathbf{b})^T \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}) - c \\ &= \frac{1}{2} (\mathbf{s} - \mathbf{b})^T \mathbf{A}^{-1}(\mathbf{s} - \mathbf{b}) - c\end{aligned}$$

which is the convex conjugate for the quadratic function.

7.11. We are using the hinge-loss (from SVM models):

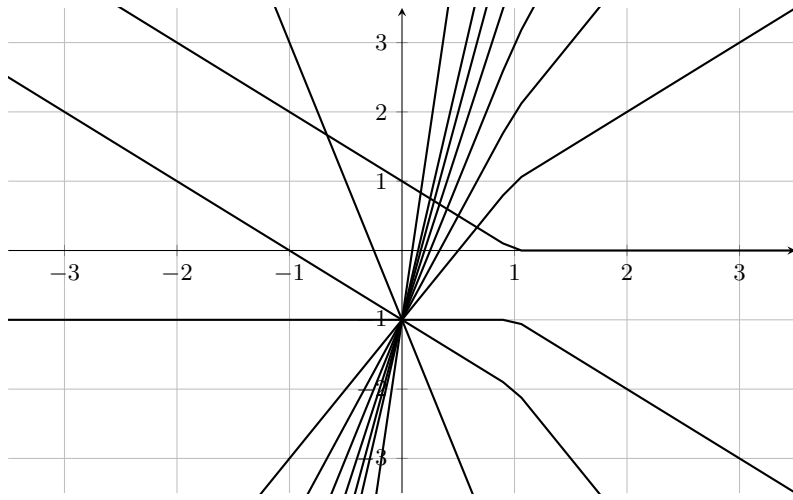
$$L(\alpha) = \max\{0, 1 - \alpha\}$$

We will calculate the convex conjugate $L^*(\beta)$. Then we will add an ℓ_2 proximal term, and calculate a second convex conjugate. (Relevant note: double conjugation always results in a *convex hull*, the smallest convex set containing a set):

$$L^*(\beta) + \frac{\gamma}{2} \beta^2$$

where γ is a given hyperparameter (so we will treat it as a constant). By the definition of the convex conjugate:

$$\begin{aligned}L^*(\beta) &= \sup_{\alpha} \{\beta\alpha - L(\alpha)\} \\ &= \sup_{\alpha} \{\beta\alpha - \max\{0, 1 - \alpha\}\} \\ &= B\end{aligned}$$



Appendix

Exercise 2.12

Code for exercise 2.12 - first plot

```
import numpy as np
import plotly.graph_objects as go

x = np.linspace(-3,3,100)
y = np.linspace(-3,3,100)

X,Y = np.meshgrid(x,y)

# Planes
Z1 = 6*X + 3*Y
Z2 = -2*X + 5*Y

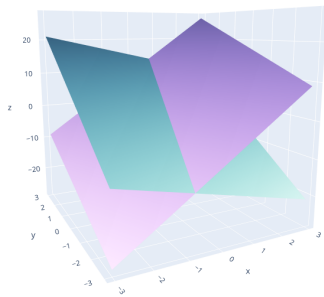
fig = go.Figure(data = [
    go.Surface(x = x,y = y,z = Z1, opacity=1.0, colorscale='Purp'),
    go.Surface(x = x,y = y,z = Z2, opacity=1.0, colorscale='Teal')
])

fig.update_layout(
    autosize=False,
    width=800,
    height=800)

fig.show()
```

Python

Results:



Exercise 2.12

Code for exercise 2.12 - second plot

```
import numpy as np
import plotly.graph_objects as go

x = np.linspace(-3,3,100)
y = np.linspace(-3,3,100)

X,Y = np.meshgrid(x,y)

# Planes
Z1 = 6*X + 3*Y
Z2 = -2*X + 5*Y

# Intersecting line
t = np.linspace(-0.75, 0.75, 500)
xl, yl, zl = t, 4*t, 18*t

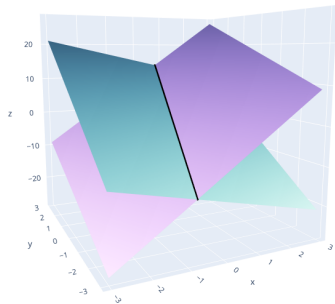
fig = go.Figure(data = [
    go.Surface(x = x,y = y,z = Z1, opacity=1.0, colorscale='Purp'),
    go.Surface(x = x,y = y,z = Z2, opacity=1.0, colorscale='Teal'),
    go.Scatter3d(x = xl, y = yl, z = zl, mode='markers',
                 marker=dict(color='rgb(0,0,0)', size=1))
])

fig.update_layout(
    autosize=False,
    width=800,
    height=800)

fig.show()
```

Python

Results:



Exercise 2.15

Code example for 2.15 c.

```
import numpy as np
import plotly.graph_objects as go

x = np.linspace(-1,1,100)
y = np.linspace(-1,1,100)
X,Y = np.meshgrid(x,y)

# Planes
Z1 = X + Y

# Axes
xaxisX = np.linspace(-2, 2, 1000)
xaxisY = 0*np.linspace(-2, 2, 1000)
xaxisZ = 0*np.linspace(-2, 2, 1000)
yaxisX = 0*np.linspace(-2, 2, 1000)
yaxisY = np.linspace(-2, 2, 1000)
yaxisZ = 0*np.linspace(-2, 2, 1000)
zaxisX = 0*np.linspace(-2, 2, 1000)
zaxisY = 0*np.linspace(-2, 2, 1000)
zaxisZ = np.linspace(-2, 2, 1000)

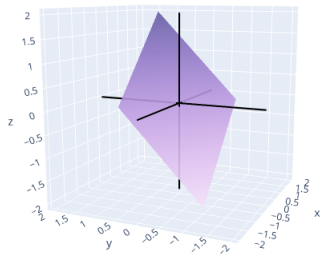
fig = go.Figure(data = [
    go.Surface(x = x,y = y,z = Z1, opacity=0.9, colorscale='Purp')
    ,go.Scatter3d(x = xaxisX, y = xaxisY, z = xaxisZ, mode='markers',
        marker=dict(color='rgb(0,0,0)', size=0.7))
    ,go.Scatter3d(x = yaxisX, y = yaxisY, z = yaxisZ, mode='markers',
        marker=dict(color='rgb(0,0,0)', size=0.7))
    ,go.Scatter3d(x = zaxisX, y = zaxisY, z = zaxisZ, mode='markers',
        marker=dict(color='rgb(0,0,0)', size=0.7))
])

fig.update_layout(
    autosize=False,
    width=800,
    height=800)

fig.show()
```

Python

Results:



Exercise 5.6

Finding the trace for the simplified calculations for exercise 5.6.

We have the following matrices:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Calculating \mathbf{AXB} .

$$\begin{aligned} \mathbf{AXB} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{11}b_{11} + x_{12}b_{21} & x_{11}b_{12} + x_{12}b_{22} \\ x_{21}b_{11} + x_{22}b_{21} & x_{21}b_{12} + x_{22}b_{22} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11}x_{11} + b_{21}x_{12} & b_{12}x_{11} + b_{22}x_{12} \\ b_{11}x_{21} + b_{21}x_{22} & b_{12}x_{21} + b_{22}x_{22} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}(b_{11}x_{11} + b_{21}x_{12}) + a_{12}(b_{11}x_{21} + b_{21}x_{22}) & a_{11}(b_{12}x_{11} + b_{22}x_{12}) + a_{12}(b_{12}x_{21} + b_{22}x_{22}) \\ a_{21}(b_{11}x_{11} + b_{21}x_{12}) + a_{22}(b_{11}x_{21} + b_{21}x_{22}) & a_{21}(b_{12}x_{11} + b_{22}x_{12}) + a_{22}(b_{12}x_{21} + b_{22}x_{22}) \end{bmatrix} \\ &= \begin{bmatrix} a_{11}b_{11}x_{11} + a_{11}b_{21}x_{12} + a_{12}b_{11}x_{21} + a_{12}b_{21}x_{22} & a_{11}b_{12}x_{11} + a_{11}b_{22}x_{12} + a_{12}b_{12}x_{21} + a_{12}b_{22}x_{22} \\ a_{21}b_{11}x_{11} + a_{21}b_{21}x_{12} + a_{22}b_{11}x_{21} + a_{22}b_{21}x_{22} & a_{21}b_{12}x_{11} + a_{21}b_{22}x_{12} + a_{22}b_{12}x_{21} + a_{22}b_{22}x_{22} \end{bmatrix} \end{aligned}$$

The trace is just the sum of the diagonal elements.

$$\begin{aligned} \text{trace}(\mathbf{AXB}) &= a_{11}b_{11}x_{11} + a_{11}b_{21}x_{12} + a_{12}b_{11}x_{21} + a_{12}b_{21}x_{22} \\ &\quad + a_{21}b_{12}x_{11} + a_{21}b_{22}x_{12} + a_{22}b_{12}x_{21} + a_{22}b_{22}x_{22} \end{aligned}$$

Exercise 6.2

R-code for making plots for exercise **6.2**. Plot number 1.

```
library(mvtnorm)
library(plotly)

DIM = 50
X1VALS = seq(from = 7, to = 12, length.out = DIM) # X
X2VALS = seq(from = -2.5, to = 3.2, length.out = DIM) # Y

N1 = length(X1VALS)
N2 = length(X2VALS)
SURF = matrix(rep(0, N1*N2), ncol = N1)

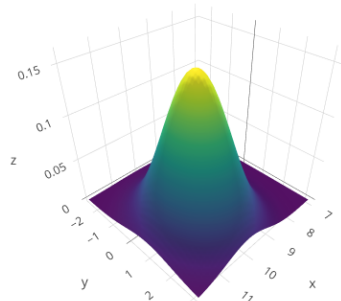
# Multivariate Gaussian parameters
means1 <- matrix(c(10,1), ncol = 1)
sigma1 <- matrix(c(1,0,0,1), ncol = 2)

for(i in 1:N1) {
  for(j in 1:N2) {
    xVal = c(X1VALS[i], X2VALS[j])
    SURF[i, j] = mvtnorm::dmvnorm(xVal,
                                  mean = means1,
                                  sigma = sigma1)
  }
}

plot_ly(x = X1VALS,
        y = X2VALS,
        z = SURF) %>% add_surface()
```

R

Results:



Exercise 6.2

Plot number 2 for exercise 6.2

```
library(mvtnorm)
library(plotly)

DIM = 50
X2VALS = seq(from = -6, to = 6, length.out = DIM) # Y
X1VALS = seq(from = -6, to = 6, length.out = DIM) # X

N1 = length(X1VALS)
N2 = length(X2VALS)
SURF = matrix(rep(0, N1*N2), ncol = N1)

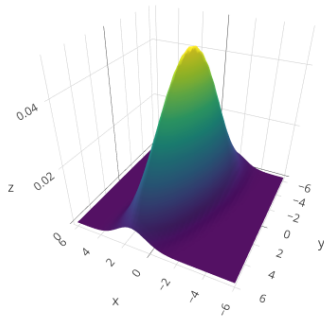
# Multivariate Gaussian parameters
means2 <- matrix(c(0,0), ncol = 1)
sigma2 <- matrix(c(8.4,2,2,1.7), ncol = 2)

for(i in 1:N1) {
  for(j in 1:N2) {
    xVal = c(X1VALS[i], X2VALS[j])
    SURF[i, j] = mvtnorm::dmvnorm(xVal,
                                  mean = means2,
                                  sigma = sigma2)
  }
}

plot_ly(x = X1VALS,
        y = X2VALS,
        z = SURF) %>% add_surface()
```

R

Results:



Exercise 6.2

Plot number 3 for exercise 6.2

```
library(mvtnorm)
library(plotly)

DIM = 50
X1VALS = seq(from = -10, to = 18, length.out = DIM) # X
X2VALS = seq(from = -4, to = 4, length.out = DIM) # Y

N1 = length(X1VALS)
N2 = length(X2VALS)
SURF = matrix(rep(0, N1*N2), ncol = N1)

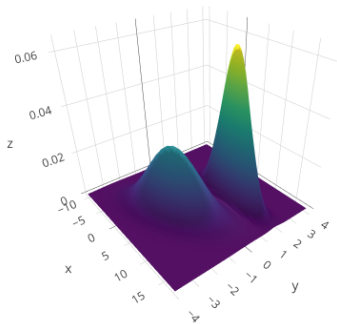
# Multivariate Gaussian parameters
means1 <- matrix(c(10,1), ncol = 1)
sigma1 <- matrix(c(1,0,0,1), ncol = 2)
means2 <- matrix(c(0,0), ncol = 1)
sigma2 <- matrix(c(8.4,2,2,1.7), ncol = 2)

for(i in 1:N1) {
  for(j in 1:N2) {
    xVal = c(X1VALS[i], X2VALS[j])
    dist1 = mvtnorm::dmvnorm(xVal,
                             mean = means1,
                             sigma = sigma1)
    dist2 = mvtnorm::dmvnorm(xVal,
                             mean = means2,
                             sigma = sigma2)
    SURF[i, j] = 0.4*dist1 + 0.6*dist2
  }
}

plot_ly(x = X1VALS,
        y = X2VALS,
        z = SURF) %>% add_surface()
```

R

Results:



Exercise 6.2

Plots for marginal distributions in exercise 6.2.

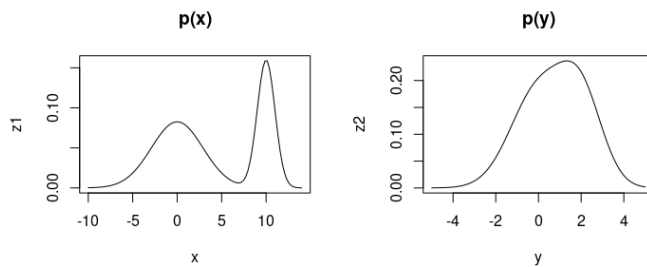
```
# Marginal distribution p(x)
x = seq(from=-10, to=14, length.out = 100)
z1 = 0.4*dnorm(x, mean = 10, sd = 1) + 0.6*dnorm(x, mean = 0, sd = sqrt(8.4))

# Marginal distribution p(y)
y = seq(from=-5, to=5, length.out = 100)
z2 = 0.4*dnorm(y, mean = 2, sd = 1) + 0.6*dnorm(y, mean = 0, sd = sqrt(1.7))

par(mfrow=c(1,2))
plot(x, z1, type="l", main = "p(x)")
plot(y, z2, type="l", main = "p(y)")
```

R

Results:



Exercise 6.2

Monte Carlo simulations for 6.2 b).

```
# Monte Carlo simulation
N = 1000000

# Calculating median
Z = rnorm(N)
zAMC1 = 0.4*(10 + 1*Z) #rnorm(N, mean = 10, sd = 1) +
zAMC2 = 0.6*(0 + sqrt(8.4)*Z) #rnorm(N, mean = 0, sd = sqrt(8.4))
M1 = median(c(zAMC1, zAMC2))
print(M1)

# Z = rnorm(N) # Makes no difference
zBMC1 = 0.4*(2 + 1*Z) #rnorm(N, mean = 10, sd = 1) +
zBMC2 = 0.6*(0 + sqrt(1.7)*Z) #rnorm(N, mean = 0, sd = sqrt(8.4))
M2 = median(c(zBMC1, zBMC2))
print(M2)

# Marginal distribution p(x)
x = seq(from=-10, to=14, length.out = 10000)
z1 = 0.4*dnorm(x, mean = 10, sd = 1) + 0.6*dnorm(x, mean = 0, sd = sqrt(8.4))

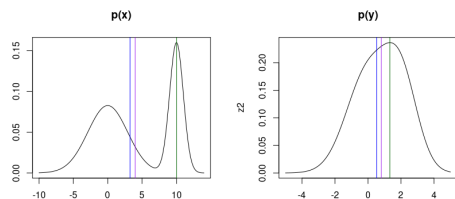
# Marginal distribution p(y)
y = seq(from=-5, to=5, length.out = 10000)
z2 = 0.4*dnorm(y, mean = 2, sd = 1) + 0.6*dnorm(y, mean = 0, sd = sqrt(1.7))

# Calculating mode
modeX = x[which.max(z1)]
modeY = y[which.max(z2)]
modeX
modeY

par(mfrow=c(1,2))
plot(x, z1, type="l", main = "p(x)")
abline(v=4, col="purple") # Mean
abline(v=3.253123, col="blue") # Median
abline(v=9.9988, col="darkgreen") # Mode
plot(y, z2, type="l", main = "p(y)")
abline(v=0.8, col="purple") # Mean
abline(v=0.5290558, col="blue") # Median
abline(v=1.330633, col="darkgreen") # Mode
```

R

Results:



Exercise 6.10

Some details on the matrix operations in exercise 6.10 b). Let's verify that it works for a 2D multivariate Gaussian to understand the relationship. Let's set $D = 2$ and:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{ss} \\ \sigma_{ss} & \sigma_{22} \end{bmatrix}$$

As mentioned in the text, we have the following for the general exponential case:

$$h(\mathbf{x}) = (2\pi)^{-\frac{D}{2}}, \quad \phi(\mathbf{x}) = \begin{bmatrix} \mathbf{x} \\ \text{vec}(\mathbf{x}\mathbf{x}^T) \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ \text{vec}(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}) \end{bmatrix}, \quad A(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{1}{2}\log|\boldsymbol{\Sigma}|$$

Also as mentioned, the vec operator simply stacks the columns in a matrix. In details::

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1x_2 \\ x_1x_2 & x_2^2 \end{bmatrix} \implies \text{vec}(\mathbf{x}\mathbf{x}^T) = \begin{bmatrix} x_1^2 \\ x_1x_2 \\ x_1x_2 \\ x_2^2 \end{bmatrix}$$

$$-\frac{1}{2}\boldsymbol{\Sigma}^{-1} = -\frac{1}{2} \begin{bmatrix} \sigma_{11} & \sigma_{ss} \\ \sigma_{ss} & \sigma_{22} \end{bmatrix} \implies \text{vec}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\right) = -\frac{1}{2} \begin{bmatrix} \sigma_{11} \\ \sigma_{ss} \\ \sigma_{ss} \\ \sigma_{22} \end{bmatrix}$$

The general exponential form can be written as:

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = h(\mathbf{x}) \exp \left[\phi(\mathbf{x})^T \boldsymbol{\theta} - A(\boldsymbol{\theta}) \right]$$

$$= (2\pi)^{-1} \exp \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_1x_2 \\ x_2^2 \end{bmatrix}^T \begin{bmatrix} \sigma_{11}\mu_1 + \sigma_{ss}\mu_2 \\ \sigma_{ss}\mu_1 + \sigma_{22}\mu_2 \\ -(1/2)\sigma_{11} \\ -(1/2)\sigma_{ss} \\ -(1/2)\sigma_{ss} \\ -(1/2)\sigma_{22} \end{bmatrix} - \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\log|\boldsymbol{\Sigma}| \right\}$$

Moving out the $\boldsymbol{\Sigma}$ term.

$$= (2\pi)^{-1} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_1x_2 \\ x_2^2 \end{bmatrix}^T \begin{bmatrix} \sigma_{11}\mu_1 + \sigma_{ss}\mu_2 \\ \sigma_{ss}\mu_1 + \sigma_{22}\mu_2 \\ -(1/2)\sigma_{11} \\ -(1/2)\sigma_{ss} \\ -(1/2)\sigma_{ss} \\ -(1/2)\sigma_{22} \end{bmatrix} - \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \right\}$$

Now some intermediary calculations. From the dot product, the first 2 terms will give:

$$x_1(\sigma_{11}\mu_1 + \sigma_{ss}\mu_2) + x_2(\sigma_{ss}\mu_1 + \sigma_{22}\mu_2) = \begin{bmatrix} \sigma_{11}\mu_1 + \sigma_{ss}\mu_2 & \sigma_{ss}\mu_1 + \sigma_{22}\mu_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{ss} \\ \sigma_{ss} & \sigma_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}$$

The last 4 elements in the dot product have a similar argument:

$$\begin{aligned}
-\frac{1}{2} \left(x_1^2 \sigma_{11} + x_1 x_2 \sigma_{ss} + x_1 x_2 \sigma_{ss} + x_2^2 \sigma_{22} \right) &= -\frac{1}{2} \begin{bmatrix} x_1 \sigma_{11} + x_2 \sigma_{ss} & x_1 \sigma_{ss} + x_2 \sigma_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= -\frac{1}{2} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{ss} \\ \sigma_{ss} & \sigma_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= -\frac{1}{2} \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x}
\end{aligned}$$

One other rewrite we will need, which works because $\boldsymbol{\Sigma}^{-1}$ is symmetric:

$$\begin{aligned}
\mathbf{x}^T \boldsymbol{\Sigma} \boldsymbol{\mu} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{ss} \\ \sigma_{ss} & \sigma_{22} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \\
&= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} \mu_1 + \sigma_{ss} \mu_2 \\ \sigma_{ss} \mu_1 + \sigma_{22} \mu_2 \end{bmatrix} \\
&= x_1 \sigma_{11} \mu_1 + x_1 \sigma_{ss} \mu_2 + x_2 \sigma_{ss} \mu_1 + x_2 \sigma_{22} \mu_2 \\
&= \mu_1 \sigma_{11} x_1 + \mu_1 \sigma_{ss} x_2 + \mu_2 \sigma_{ss} x_1 + \mu_2 \sigma_{22} x_2 \\
&= \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} x_1 + \sigma_{ss} x_2 \\ \sigma_{ss} x_1 + \sigma_{22} x_2 \end{bmatrix} \\
&= \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{ss} \\ \sigma_{ss} & \sigma_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= \boldsymbol{\mu}^T \boldsymbol{\Sigma} \mathbf{x}
\end{aligned}$$

Returning to the main calculation and making the replacements.

$$\begin{aligned}
f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= (2\pi)^{-1} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_1 x_2 \\ x_2^2 \end{bmatrix}^T \begin{bmatrix} \sigma_{11} \mu_1 + \sigma_{ss} \mu_2 \\ \sigma_{ss} \mu_1 + \sigma_{22} \mu_2 \\ -(1/2) \sigma_{11} \\ -(1/2) \sigma_{ss} \\ -(1/2) \sigma_{ss} \\ -(1/2) \sigma_{22} \end{bmatrix} - \frac{1}{2} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right\} \\
&= (2\pi)^{-1} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} + \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right\} \\
&= (2\pi)^{-1} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - 2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} + \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right] \right\} \\
&= (2\pi)^{-1} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} + \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right] \right\}
\end{aligned}$$

Rewriting term as shown above.

$$\begin{aligned}
&= (2\pi)^{-1} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} + \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \right] \right\} \\
&= (2\pi)^{-1} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} \sim \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})
\end{aligned}$$

Which verifies the relationship! For further details, see e.g. this blog post:

<https://maurocamaraescudero.netlify.app/post/multivariate-normal-as-an-exponential-family-distribution/>

Exercise 7.6

PRIMAL PROBLEM for exercise 7.6.

```
import numpy as np
from scipy.optimize import linprog

A = np.array([[2, 2], [2,-4], [-2, 1], [0, -1], [0, 1]])
b = np.array([33,8,5,-1,8])
c = np.array([-5,-3])

res = linprog(c, A_ub=A, b_ub=b, bounds=(None, None))

print("Optimal value:", res.fun.round(3), "\nX:", res.x.round(3))
```

Python

Output:

```
Optimal value: -74.167
X: [12.333  4.167]
```

DUAL PROBLEM for exercise 7.6.

```
import numpy as np
from scipy.optimize import linprog

A = np.array([[2, 2, -2, 0, 0], [2,-4, 1, -1, 1]])
b = np.array([5,3])
c = np.array([33,8,5,-1,8]) # max [-b^T]*lambda corresponds to min [b^T]*lambda

res = linprog(c, A_eq=A, b_eq=b, bounds=(0, None))

print("Optimal value:", res.fun.round(3), "\nLambda:", res.x.round(3))
```

Python

Output:

```
Optimal value: 74.167
Lambda: [2.167 0.333 0.  0.  0.  ]
```

Comment: We have to solve this as a minimum optimization problem in Python, so we switch $-\mathbf{b}$ to \mathbf{b} in the code. Gives the correct lambda values, but the optimal value gets the incorrect sign.