



BME

LECTURE NOTES FOR THE SUBJECT

Mathematics MSc for Civil Engineers

Levente David

lectured by
Balázs Bárány

September 24, 2025

Preface

This is an extended lecture note for the subjects "Mathematics MSc for Civil Engineers" and "Construction Information Technology Mathematics" taught at the Budapest University of Technology and Economics in English language.

The project supported by the Doctoral Excellence Fellowship Programme (KCEP) is funded by the National Research, Development and Innovation Fund of the Ministry of Culture and Innovation and the Budapest University of Technology and Economics.

Contents

1	Linear algebra	4
1.1	Matrices	4
1.1.1	Basic operations	4
1.1.2	Determinant	6
1.1.3	Exercises	9
1.2	Linear Equations and Linear Equation Systems	9
1.2.1	Exercises	16
1.3	Vector spaces	16
1.3.1	Exercises	19
1.4	Linear independence, basis, basis representation	20
1.4.1	Finding a basis of the spanned subspace	23
1.4.2	Equivalent conditions for invertibility	24
1.4.3	Change of basis	25
1.4.4	Exercises	27
1.5	Linear transformations	27
1.5.1	Exercises	30
1.6	Eigenvalues, eigenvectors	30
1.6.1	Exercises	33
1.7	Scalar product	33
1.8	Symmetric matrices	35
1.8.1	Gram-Schmidt orthogonalisation	36
1.8.2	Quadratic Forms	38
1.8.3	Exercises	42
1.9	Trace and double dot product*	42
1.9.1	Exercises	44
1.10	Fundamental Subspaces of a Matrix	44
1.10.1	Orthogonal complements	47
1.10.2	Exercises	50
1.11	Orthogonal projections	50
1.11.1	Method of least squares	54
1.11.2	Exercises	57
1.12	Decompositions of matrices	57
1.12.1	Definite matrices	57

1.12.2 Singular value decomposition (SVD)	59
1.12.3 Spectral decomposition	62
1.12.4 Exercises	65
1.13 Mixed exercises in Linear Algebra	65

Chapter 1

Linear algebra

1.1 Matrices

We begin by recalling some basic definitions from our previous studies on linear algebra. First, let us consider one of the basic objects, namely, the matrices.

Let m and n be positive integers. The matrix $m \times n$ $A = [a_{ij}] = [a_{ij}]_{i=1,j=1}^{m,n}$ is a table of numbers

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

with m rows and n columns, where $a_{i,j}$ are reals indexed by two index set: $i \in \{1, \dots, m\}, j \in \{1, \dots, n\}$. We denote the set of $m \times n$ matrices by $\mathbb{R}^{m \times n}$.

1.1.1 Basic operations

For given two $m \times n$ matrices $A = [a_{ij}] \in \mathbb{R}^{m \times n}$ and $B = [b_{ij}] \in \mathbb{R}^{m \times n}$, we define the sum $A + B$ as an $m \times n$ matrix of which element in the i th row and j th column is $a_{ij} + b_{ij}$. That is,

$$A+B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}.$$

Similarly, the multiplication of a matrix $A = [a_{ij}] \in \mathbb{R}^{m \times n}$ with a scalar $c \in \mathbb{R}$, we define the matrix cA as an $m \times n$ matrix of which element in the i th row and j th column is ca_{ij} . That is,

$$cA = c \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{bmatrix}.$$

Given two matrices, $A = [a_{ij}] \in \mathbb{R}^{m \times n}$ and $B = [b_{k\ell}] \in \mathbb{R}^{n \times h}$ we can define $A \cdot B \in \mathbb{R}^{m \times h}$, the product of matrices, with the equation:

$$AB = \left[\sum_{j=1}^n a_{ij} b_{j\ell} \right]_{i=1, \ell=1}^{m, h}.$$

That is, the AB matrix has m rows and h columns and its i, ℓ th element is $\sum_{j=1}^n a_{ij} b_{j\ell}$.

Example 1.1.1. Let us multiply a 2×3 matrix with a 3×4 matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 2 & 1 \\ -1 & 3 & 1 & 0 \\ 2 & 4 & 0 & -2 \end{bmatrix}.$$

The product AB is defined since the inner dimensions (3 and 3) agree, and the result is a 2×4 matrix.

$$\begin{aligned} AB &= \begin{bmatrix} 1 \cdot 1 + 2 \cdot (-1) + 3 \cdot 2 & 1 \cdot 0 + 2 \cdot 3 + 3 \cdot 4 & 1 \cdot 2 + 2 \cdot 1 + 3 \cdot 0 & 1 \cdot 1 + 2 \cdot 0 + 3 \cdot (-2) \\ 4 \cdot 1 + 5 \cdot (-1) + 6 \cdot 2 & 4 \cdot 0 + 5 \cdot 3 + 6 \cdot 4 & 4 \cdot 2 + 5 \cdot 1 + 6 \cdot 0 & 4 \cdot 1 + 5 \cdot 0 + 6 \cdot (-2) \end{bmatrix} \\ &= \begin{bmatrix} 5 & 18 & 4 & -5 \\ 11 & 39 & 13 & -8 \end{bmatrix}. \end{aligned}$$

Thus AB is a 2×4 matrix.

Let us note that unlike the multiplication of real numbers, the multiplication of matrices is not commutative.

Example 1.1.2. Consider two 2×2 matrices:

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

Compute

$$AB = \begin{bmatrix} 1 \cdot 0 + 2 \cdot 1 & 1 \cdot 1 + 2 \cdot 0 \\ 0 \cdot 0 + 1 \cdot 1 & 0 \cdot 1 + 1 \cdot 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix},$$

while

$$BA = \begin{bmatrix} 0 \cdot 1 + 1 \cdot 0 & 0 \cdot 2 + 1 \cdot 1 \\ 1 \cdot 1 + 0 \cdot 0 & 1 \cdot 2 + 0 \cdot 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}.$$

Since

$$AB = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} = BA,$$

we see that matrix multiplication is generally not commutative.

Definition 1.1.3. Let $A = [a_{ij}] \in \mathbb{R}^{m \times n}$. The **transpose** of A , denoted $A^\top$, is the $n \times m$ matrix of which element in the i th row and j th column is a_{ji} . That is, the rows of A become the columns of

A^T , and vice versa. That is,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \text{then} \quad A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

Proposition 1.1.4. For $A, B \in \mathbb{R}^{m \times n}$,

$$(A + B)^T = A^T + B^T.$$

Moreover, for two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times h}$ we have that

$$(AB)^T = B^T A^T.$$

1.1.2 Determinant

Definition 1.1.5. The **determinant** is a function that assigns to every square matrix $A \in \mathbb{R}^{n \times n}$ a scalar $\det(A) \in \mathbb{R}$, defined inductively as follows:

- For $n = 1$, if $A = [a]$, then $\det(A) = a$.
- For $n = 2$, if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $\det(A) = ad - bc$.
- For $n \geq 2$, the determinant is defined by expanding along the first row:

$$\det(A) = \sum_{j=1}^n (-1)^{1+j} a_{1j} \det(A_{1j}),$$

where A_{1j} is the $(n-1) \times (n-1)$ matrix obtained by deleting the first row and j -th column of A . These submatrices are called *minors* of A . For example for a 4×4 matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \rightsquigarrow A_{12} = \begin{bmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{bmatrix}.$$

Theorem 1.1.6. Let $A \in \mathbb{R}^{n \times n}$. The determinant of A can be computed by expanding along any row or any column. Specifically, for any fixed row i or column j ,

$$\det(A) = \sum_{k=1}^n (-1)^{i+k} a_{ik} \det(A_{ik}) = \sum_{k=1}^n (-1)^{k+j} a_{kj} \det(A_{kj}),$$

where A_{ik} or A_{kj} are the appropriate minors of A .

Theorem 1.1.7. For a 3×3 matrix, the determinant can be computed using a shortcut known as

Sarrus' Rule. Let

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}.$$

Then the determinant is:

$$\det(A) = aei + bfg + cdh - ceg - bdi - afh.$$

Remark 1.1.8. Sarrus' rule can be visualized by rewriting the first two columns of the matrix next to it and then summing the products of the diagonals from upper left to lower right, and subtracting the products of the diagonals from lower left to upper right.

$$\left[\begin{array}{ccccc} a & b & c & a & b \\ d & e & f & d & e \\ g & h & i & g & h \end{array} \right]$$

The red diagonals correspond to the positive terms

$$aei + bfg + cdh,$$

while the blue diagonals correspond to the negative terms

$$ceg + bdi + afh.$$

?(ex:3b3)?

Example 1.1.9 (Determinant of a 3×3 Matrix). Find the determinant $\det(A)$ of the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

We apply Sarrus' rule. That is, we copy the first two columns to the right:

$$\left[\begin{array}{ccc|cc} 1 & 2 & 3 & 1 & 2 \\ 4 & 5 & 6 & 4 & 5 \\ 7 & 8 & 9 & 7 & 8 \end{array} \right].$$

Now compute the diagonal products:

$$(1 \cdot 5 \cdot 9) + (2 \cdot 6 \cdot 7) + (3 \cdot 4 \cdot 8) = 45 + 84 + 96 = 225,$$

and the antidiagonal products:

$$(3 \cdot 5 \cdot 7) + (1 \cdot 6 \cdot 8) + (2 \cdot 4 \cdot 9) = 105 + 48 + 72 = 225.$$

Subtracting:

$$\det(A) = 225 - 225 = 0.$$

?<prop:det>?

Theorem 1.1.10. Let $A, B \in \mathbb{R}^{n \times n}$, and let $c \in \mathbb{R}$. Then:

1. $\det(A^\top) = \det(A)$,
2. $\det(cA) = c^n \det(A)$,
3. $\det(AB) = \det(A) \det(B)$.

Remark 1.1.11. In $\mathbb{R}^2$, the determinant gives the signed area of the parallelogram spanned by two column vectors of A . In $\mathbb{R}^3$, it gives the signed volume of the parallelepiped defined by the columns of A . See also later, Remark 1.5.7.

Definition 1.1.12 (Inverse of a Matrix). Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. If there exists a matrix $B \in \mathbb{R}^{n \times n}$ such that

$$AB = BA = I_n,$$

where

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

is the $n \times n$ identity matrix, then A is said to be **invertible**, and B is called the **inverse** of A , denoted A^{-1} .

Theorem 1.1.13 (Formulas for case 2×2). For a 2×2 matrix

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

if $\det(A) = ad - bc \neq 0$, then

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Proof. Let us check the product

$$\frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} da - cb & -ab + ba \\ -ac + ca & -bc + ad \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

□

Example 1.1.14 (Inverse of a 2×2 Matrix). Let

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

First compute the determinant:

$$\det(A) = 2 \cdot 3 - 1 \cdot 5 = 6 - 5 = 1.$$

Since $\det(A) \neq 0$, A is invertible. By the formula,

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

Theorem 1.1.15. A square matrix $A \in \mathbb{R}^{n \times n}$ is invertible if and only if $\det(A) \neq 0$.

Example 1.1.16. To see a non-invertible matrix, use the previous corollary to see that the matrix in Example 1.1.9 is not invertible.

1.1.3 Exercises

1. Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 2 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 2 & 3 \\ 1 & 4 & 2 \\ 3 & 2 & 1 \end{bmatrix}.$$

Calculate $A(BC)$ and $(AB)C$.

2. Calculate the determinant of

$$A = \begin{bmatrix} 11 & -2 & -1 \\ 3 & 4 & -8 \\ 7 & 1 & 2 \end{bmatrix}.$$

3. Calculate the inverse of

$$A = \begin{bmatrix} 61 & 24 \\ 29 & 4 \end{bmatrix}.$$

1.2 Linear Equations and Linear Equation Systems

Definition 1.2.1. A **linear equation** of n variables $x_1, x_2, \dots, x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

where $a_1, a_2, \dots, a_n$ and b are real (or complex) numbers, and $x_1, x_2, \dots, x_n$ are variables, that is, unknown numbers. A **system of linear equations** is a collection of m linear equations of n variables:

$$\left. \begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \right\} \quad (1.2.1) \text{ ?eq:lineq?}$$

We call a system of linear equations **homogeneous** if $b_1 = \cdots = b_m = 0$.

The central goal is to find all tuples $(x_1, \dots, x_n)$ that satisfy every equation in the system simultaneously. Let us note that a homogeneous linear equation system always has (at least one) solution by taking $x_1 = \dots = x_n = 0$. To find solutions, we are allowed to perform a set of operations that preserve the solution set. These are known as the **elementary operations**:

- Exchange the positions of two equations.
- Multiply an equation by a non-zero scalar.
- Replace one equation with the sum of itself and a scalar multiple of another equation.

These operations clearly do not change the solution. Using such operations, we aim to transform a given system into a simpler one, from which the solution is easier to determine. First, let us introduce two notations.

Definition 1.2.2. Given a system of m linear equations in n variables as in (1.2.1), the **coefficient matrix** A is the $m \times n$ matrix containing the coefficients of the variables:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

The **extended coefficient matrix** (or simply, **augmented matrix**) of the system is the $m \times (n + 1)$ matrix obtained by appending the column of constants to the coefficient matrix:

$$[A \mid \mathbf{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right].$$

Using the coefficient matrix A , and defining $\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n \times 1}$ and $\underline{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^{m \times 1}$, we

can write (1.2.1) in the form

$$A\underline{x} = \underline{b}. \tag{1.2.2}$$

For every $i = 1, \dots, n$, the i th column of the augmented matrix corresponds to the variable x_i by representing its coefficients in the linear equation system. Each elementary operation of a linear equation system corresponds naturally to the manipulations of rows in the augmented matrix. These operations are called **elementary row manipulations**:

- Exchange the position of two rows in the augmented matrix.
- Multiply a row by a non-zero scalar.
- Replace a row with the sum of itself and scalar multiple of another row.

Let us see now an example to demonstrate how to use elementary row manipulations for solving linear equation systems.

Example 1.2.3. Consider the following system of linear equations of four variables:

$$\left. \begin{aligned} x_1 + 2x_2 - x_3 + x_4 &= 3, \\ 2x_1 + 5x_2 + x_3 + 4x_4 &= 10, \\ 3x_1 + 7x_2 + 5x_4 &= 13. \end{aligned} \right\} \quad (1.2.3) \text{ ?eq:ex2?}$$

Then We write its extended matrix:

$$\left[\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 3 \\ 2 & 5 & 1 & 4 & 10 \\ 3 & 7 & 0 & 5 & 13 \end{array} \right].$$

The general idea in solving linear equations is as follows: the easiest way to read the solution would be if every variable would appear in the linear equation in a "reversed chronological" order (i.e. the first variable appears in the first the first equation, but not in the others; the second appears in the first and second but nowhere else etc.). This is not always possible, but it is a good guideline for our strategy. Along these lines, let us do the following:

Step 1: Eliminate entries below 1 in the first column. First, subtract the double of row 1 from row 2:

$$R_2 \leftarrow R_2 - 2R_1 : \left[\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 3 \\ 0 & 1 & 3 & 2 & 4 \\ 3 & 7 & 0 & 5 & 13 \end{array} \right].$$

Then subtract the triple of row 1 from row 3:

$$R_3 \leftarrow R_3 - 3R_1 : \left[\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 3 \\ 0 & 1 & 3 & 2 & 4 \\ 0 & 1 & 3 & 2 & 4 \end{array} \right].$$

As we see, every element in the first column is zero except the first one. In particular, we eliminated the first variable from every equation except the first, where its coefficient is 1. Now, we intend to do the same with the second variable.

Step 2: Eliminate entries below the first non-zero element in the second column: Subtract row 2 from row 3:

$$R_3 \leftarrow R_3 - R_2 : \left[\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 3 \\ 0 & 1 & 3 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]. \quad (1.2.4) \text{ ?eq:ex1?}$$

This is a stage where we can determine the solution. The augmented matrix above corresponds to the following linear equation system:

$$\begin{aligned} x_1 + 2x_2 - x_3 + x_4 &= 3, \\ x_2 + 3x_3 + 2x_4 &= 4, \\ 0 &= 0. \end{aligned}$$

The last row corresponds to the trivial equation $0 = 0$ therefore we can omit it. Since there are four variables and only two non-trivial equations, we can choose two variables freely. Set $x_3 = s$, $x_4 = t$, where $s, t \in \mathbb{R}$. Then we can express x_2 as

$$x_2 + 3s + 2t = 4 \quad \Rightarrow \quad x_2 = 4 - 3s - 2t.$$

Now, using this, we can also express x_1 by s and t

$$x_1 + 2x_2 - s + t = 3.$$

Substitute $x_2 = 4 - 3s - 2t$:

$$x_1 + 2(4 - 3s - 2t) - s + t = 3 \Rightarrow x_1 + 8 - 6s - 4t - s + t = 3 \Rightarrow x_1 = 3 - 8 + 7s + 3t = -5 + 7s + 3t.$$

So, the solution is:

$$x_1 = -5 + 7s + 3t,$$

$$x_2 = 4 - 3s - 2t,$$

$$x_3 = s,$$

$$x_4 = t,$$

where $s, t \in \mathbb{R}$ can be chosen freely, that is, the choice of $x_1, \dots, x_4$ above solves the linear equation system in (1.2.3) for every values of s and t .

This has augmented matrix

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

Notice, that this last step could be easily done by working only on the augmented matrix by subtracting the double of row 2 from row 1.

There is a general strategy for solving linear equations in the form (1.2.1), which is called the **Gauss-Jordan elimination**. It is an algorithmic procedure for solving systems of linear equations by systematically applying **elementary row operations** to transform the augmented matrix of a system into a special form known as the **reduced row echelon form** (RREF).

?<def:RREF>?

Definition 1.2.4. We say that an $m \times n$ matrix A has **row echelon form** (REF) if

- all rows containing non-zero elements are above any rows which contain only zeros;
- the leftmost non-zero entry in each non-zero row is 1. This is called a **pivot element**;
- in any two consecutive rows, the pivot element of the row below stands strictly to the right of the pivot of the row above.

We say that an $m \times n$ matrix A has **reduced row echelon form** (RREF) if

- it has a row echelon form;
- Entries **above and below** each pivot are zero.

Clearly, in a matrix of row echelon form, the entries below each pivot elements are zero.

Example 1.2.5. The following matrices are in row-echelon form:

$$\begin{bmatrix} 1 & 4 & 3 & 7 \\ 0 & 1 & 6 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}; \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

The following matrices are in reduced row-echelon form:

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

Algorithm 1.2.6 (Gauss-Jordan elimination). Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

?(st:1)?

- (i) If the first column of A contains only zeros then replace cover the first column and continue the procedure with the uncovered part of the matrix from (i). Otherwise go to the next step.
- (ii) If the first column of A contains a non-zero element then exchange the first row and the row with non-zero first element. If the first element of the first column is non-zero, then go to next step.
- (iii) Divide the first row by its first (non-zero) element. So we are at stage:

$$\begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- (iv) Subtract from the lower rows the corresponding multiplier of the first row to make the first elements in the lower rows zero. That is, subtract from row k the a_{k1} multiplier of the row 1. So we are at stage:

$$\begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- (v) Cover the first row and first column of the matrix. If there is nothing left, stop, otherwise continue the process on the uncovered part from (i).

The matrix at the end of the process described above has row-echelon form. Then continue with the following:

?(st:1b)?

- (vi) If the last column of A contains only zeros then replace cover the last column and continue the procedure with the uncovered part of the matrix from (vi).
- (vii) If the last column of the matrix has a non-zero element then the last non-zero element of the column must be a pivot. Then subtract from the upper rows the corresponding multiplier of that row to make all the elements above zero. Then cover the last column of the matrix. If nothing left then stop, otherwise go to (vi).

Let us note that we need to do every manipulation described above with the **covered parts** too!

The solution can be easily read from the reduced row-echelon form if we rewrite it in the linear equation system form.

rop:solutions)?

Proposition 1.2.7. The linear equation system

$$\left. \begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \right\}$$

- has no solution, if the reduced row-echelon form of its augmented matrix (after the Gauss-Jordan elimination) contains a pivot element in the last column;
- has a unique solution if every column in the reduced row-echelon form of its augmented matrix (after the Gauss-Jordan elimination) contains a pivot element except the last column;
- has infinitely many solutions, if the last column and another column in the reduced row-echelon form of its augmented matrix (after the Gauss-Jordan elimination) do not contain pivot element. In that case, the free parameters can be chosen to be the variables of which corresponding column does not contain pivot elements.

Proof. Since we know that the elementary row manipulations do not change the solution, it is enough to focus on linear equation systems with augmented matrix in reduced row-echelon form. Let A be such augmented matrix. If the last column contains a pivot element then it has the form

$$\left[\begin{array}{cccc|c} 1 & a_{12} & \cdots & a_{1n} & b_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & b_m \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right].$$

Rewriting the augmented matrix in the linear equation system form, the last non-zero row gives the equation $0 = 1$, which is absurd, so the equation system does not have a solution.

If every column of the reduced row-echelon form contains a pivot element except the last one, then it must have the form

$$\left[\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & b_1 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & b_n \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right].$$

Rewriting the augmented matrix in the linear equation system form, we get that $x_1 = b_1, \dots, x_n = b_n$, which is the unique solution.

We can solve the remaining case, when the last column and another column do not contain pivot elements, then rewriting the augmented matrix back to the linear equation form, we can express every variable which corresponds to a column with pivot element using variables corresponding to columns without pivot elements. These latter ones are called free variables. $\square$

Let us demonstrate the Gauss-Jordan elimination on the following example.

Example 1.2.8. Solve the following linear equation system:

$$\left. \begin{array}{l} x + 2y - z = 3 \\ 2x + 4y + z = 7 \\ 2z - x = 0. \end{array} \right\}$$

Let us rewrite it into the augmented matrix form.

$$\begin{aligned} \left[\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 2 & 4 & 1 & 7 \\ -1 & 0 & 2 & 0 \end{array} \right] &\xrightarrow{\substack{R_2 \leftarrow R_2 - 2R_1, \\ R_3 \leftarrow R_3 + R_1}} \left[\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & 0 & 3 & 1 \\ 0 & 2 & 1 & 3 \end{array} \right] &\xrightarrow{R_2 \leftrightarrow R_3} \left[\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & 2 & 1 & 3 \\ 0 & 0 & 3 & 1 \end{array} \right] \\ &\xrightarrow{R_2 \leftarrow R_2/3} \left[\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 3 & 1 \end{array} \right] &\xrightarrow{R_3 \leftarrow \frac{1}{3}R_3} \left[\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 1 & 1/3 \end{array} \right] \\ &\xrightarrow{\substack{R_2 \leftarrow R_2 - \frac{1}{2}R_3 \\ R_1 \leftarrow R_1 + R_3}} \left[\begin{array}{ccc|c} 1 & 2 & 0 & \frac{10}{3} \\ 0 & 1 & 0 & \frac{4}{3} \\ 0 & 0 & 1 & \frac{1}{3} \end{array} \right] &\xrightarrow{R_1 \leftarrow R_1 - 2R_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{2}{3} \\ 0 & 1 & 0 & \frac{4}{3} \\ 0 & 0 & 1 & \frac{1}{3} \end{array} \right] \end{aligned}$$

That is, the linear equation system has the unique solution $x = \frac{2}{3}$, $y = \frac{4}{3}$ and $z = \frac{1}{3}$.

The Gauss-Jordan elimination is one of the most important tools in linear algebra. Solving linear equation systems is not only important on its own but many problems can be reduced to solving a system of linear equations.

1.2.1 Exercises

1. Solve the following homogeneous linear equation system

$$\left. \begin{aligned} x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 &= 0, \\ -x_1 + x_2 - 6x_4 + x_5 &= 0, \\ 3x_2 + x_3 + 5x_4 - x_5 &= 0 \\ 2x_3 - 7x_4 + 7x_5 &= 0. \end{aligned} \right\}$$

2. Let us consider the following system of linear equations

$$\left. \begin{aligned} 2x - y + z &= 4, \\ 2y + 3z &= 3, \\ x + y + az &= b. \end{aligned} \right\}$$

How shall we choose the parameters a and b such that

- (a) the system has no solution,
- (b) the system has a unique solution,
- (c) the system has infinitely many solutions.

In all of the cases (when it is possible), give all solutions!

1.3 Vector spaces

We now continue with the basic object of linear algebra, the vector spaces.

Definition 1.3.1. A **vector space** over the reals $\mathbb{R}$ is a set V equipped with two operations vector addition and scalar multiplication such that for all $\underline{u}, \underline{v}, \underline{w} \in V$ and all scalars $a, b \in \mathbb{R}$, the following axioms hold:

1. $\underline{u} + \underline{v} = \underline{v} + \underline{u}$ (Commutativity of $+$)
2. $(\underline{u} + \underline{v}) + \underline{w} = \underline{u} + (\underline{v} + \underline{w})$ (Associativity of $+$)
3. There exists a zero vector $\underline{0} \in V$ such that $\underline{u} + \underline{0} = \underline{u}$ (Zero element)
4. For each $\underline{u} \in V$, there exists $-\underline{u} \in V$ such that $\underline{u} + (-\underline{u}) = \underline{0}$ (Inverse)
5. $a(\underline{u} + \underline{v}) = a\underline{u} + a\underline{v}$ and $(a + b)\underline{u} = a\underline{u} + b\underline{u}$ (Distributive properties)
6. $1 \cdot \underline{u} = \underline{u}$ and $0 \cdot \underline{u} = \underline{0}$.

Example 1.3.2. Another natural example of vector spaces is the space of row vectors $\mathbb{R}^n$. That is, let $n \in \mathbb{N}$ be positive integer and let us write

$$\mathbb{R}^n = \{(x_1, \dots, x_n) : x_1, \dots, x_n \in \mathbb{R}\}.$$

For every $\underline{x} = (x_1, \dots, x_n), \underline{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$ and $a \in \mathbb{R}$, let us define the following operations

$$\underline{x} + \underline{y} := (x_1 + y_1, \dots, x_n + y_n) \text{ and } a \cdot \underline{x} := (a \cdot x_1, \dots, a \cdot x_n).$$

With these operations, $\mathbb{R}^n$ is a vectorspace. Usually, the vectors described in Example ?? are represented by the vectors in $\mathbb{R}^3$.

Example 1.3.3 (Vector addition in $\mathbb{R}^2$). Let $\underline{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\underline{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ be vectors in $\mathbb{R}^2$. Vector addition is defined componentwise:

$$\underline{u} + \underline{v} = \begin{bmatrix} 2 + 1 \\ 1 + 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

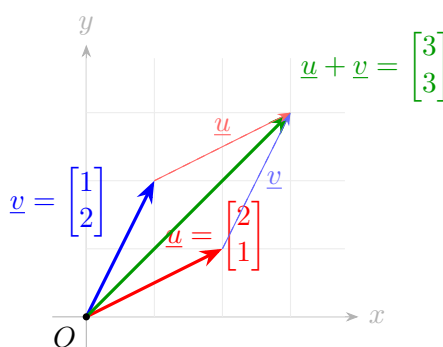


Figure 1.1: Geometric interpretation of vector addition. The sum $\underline{u} + \underline{v}$ can be obtained by placing the tail of $\underline{v}$ at the tip of $\underline{u}$; equivalently, $\underline{u} + \underline{v}$ is the diagonal of the parallelogram spanned by $\underline{u}$ and $\underline{v}$. The red arrow is $\underline{u}$, the blue arrow is $\underline{v}$, the blue arrow translated to the tip of $\underline{u}$ is drawn dashed, and the green arrow is $\underline{u} + \underline{v}$.

With a slight abuse of notation, we will also use the notation $\mathbb{R}^n$ for the vector space of column vectors

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1, \dots, x_n \in \mathbb{R} \right\}.$$

The operations are defined similarly, that is, we take the sum and scalar product element-wise.

Definition 1.3.4. Let $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n$ be vectors in a vector space V over $\mathbb{R}$. A **linear combination** of these vectors is any vector of the form

$$a_1 \underline{v}_1 + a_2 \underline{v}_2 + \dots + a_n \underline{v}_n,$$

where $a_1, a_2, \dots, a_n \in \mathbb{R}$. The scalars are called the **coefficients** of the linear combination.

?<ex2>?

Example 1.3.5. Consider the linear equation system $\left. \begin{array}{l} x_1 - x_2 = -1 \\ 5x_1 + 2x_2 = 8 \end{array} \right\}$. Then the linear equation

can be written in the form

$$x_1 \begin{bmatrix} 1 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 8 \end{bmatrix}.$$

This means solving linear equation systems is equivalent to finding coefficients x_1, x_2 such that the corresponding linear combination of the vectors $\underline{v}_1$ and $\underline{v}_2$ equals to $\underline{b}$, where $\underline{v}_1 := \begin{bmatrix} 1 \\ 5 \end{bmatrix}$, $\underline{v}_2 := \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, $\underline{b} := \begin{bmatrix} -1 \\ 8 \end{bmatrix}$.

Example 1.3.5 shows actually a general phenomenon. Consider the linear equation system

$$\left. \begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \right\} \quad (1.3.1) \text{ ?eq:lineq2?}$$

Let

$$\underline{v}_i := \begin{bmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{bmatrix} \in \mathbb{R}^m \text{ for every } i = 1, \dots, n, \text{ and let } \underline{b} := \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m. \quad (1.3.2) \text{ ?eq:coeffvec?}$$

Then the linear equation system (1.3.1) can be written in the form

$$x_1 \underline{v}_1 + \cdots + x_n \underline{v}_n = \underline{b}.$$

Hence, solving the linear equation system 1.3.1 is equivalent with the question: Does there exists coefficients $x_1, \dots, x_n$ such that $\underline{b}$ can be expressed as the linear combination of the vectors $\underline{v}_1, \dots, \underline{v}_n$ (which are column vectors of the coefficient matrix of the linear equation system). Proposition 1.2.7 provides us a method to answer this question and now, we will discuss what consequences does it have for vector spaces.

Definition 1.3.6. Let $W \subseteq V$ be a subset of the vector space V over $\mathbb{R}$. We call W a **linear subspace** if:

- The zero vector $\underline{0} \in W$;
- W is closed under vector addition, that is, $\underline{u} + \underline{v} \in W$ for every $\underline{u}, \underline{v} \in W$;
- W is closed under scalar multiplication: $a\underline{v} \in W$ for every $a \in \mathbb{R}$ and $\underline{v} \in W$.

In other words, a subspace is a subset of V that is itself a vector space under the same operations.

Example 1.3.7. The sets

$$K = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 : x + y + z = 0 \right\}, \quad \left\{ \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \in \mathbb{R}^3 : x, y \in \mathbb{R} \right\}$$

are subspaces of $\mathbb{R}^3$. For example consider two vectors $\underline{a}_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$ and $\underline{a}_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \in K$. Now their

sum, $\underline{a}_1 + \underline{a}_2 = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{bmatrix}$, and clearly

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

Hence, $\underline{a}_1 + \underline{a}_2 \in K$. Moreover, for a constant s , we have that $s\underline{a}_1 = \begin{bmatrix} sx_1 \\ sy_1 \\ sz_1 \end{bmatrix}$, and

$$sx_1 + sy_1 + sz_1 = s(x_1 + y_1 + z_1) = s \cdot 0 = 0.$$

Thus, $s\underline{a}_1 \in K$.

Example 1.3.8 (Not a Subspace). The sets

$$\left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 : x + y + z = 1 \right\}, \quad \left\{ \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \in \mathbb{R}^3 : x, y \in \mathbb{R} \right\}$$

are not subspaces of $\mathbb{R}^3$, since they do not contain the origin.

?{def:span}?

Definition 1.3.9. Given vectors $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n \in V$. We call the set of all possible linear combinations of these vectors by the subspace **spanned** by the vectors $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n$, and we denote it as

$$\text{span}(\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n) = \{a_1\underline{v}_1 + a_2\underline{v}_2 + \dots + a_n\underline{v}_n : a_i \in \mathbb{R}\}.$$

The span is always a subspace of V .

Using the definition of spanned subspace, it is easy to see that the linear equation system (1.3.1) has a solution (infinitely many or a unique) if and only if

$$\underline{b} \in \text{span}(\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n),$$

where the vectors $\underline{b}$ and $\underline{v}_i$ are as in (1.3.2). To decide whether the solution is unique or not, we need a better understanding on the vectors $\underline{v}_1, \dots, \underline{v}_n$.

1.3.1 Exercises

1. Let V be a vector space and let $\underline{v}_1, \dots, \underline{v}_n \in V$. Show that $\text{span}(\underline{v}_1, \dots, \underline{v}_n)$ is a subspace of V .

1.4 Linear independence, basis, basis representation

Definition 1.4.1. Let V be a vector space. The vectors $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n$ in V are said to be **linearly independent** if the equation

$$a_1 \underline{v}_1 + a_2 \underline{v}_2 + \dots + a_n \underline{v}_n = \underline{0}$$

has only the trivial solution: $a_1 = a_2 = \dots = a_n = 0$. If there exist non-trivial scalars $a_1, \dots, a_n$ satisfying the equation, the vectors are called **linearly dependent**.

Definition 1.4.2. A set of vectors $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n \in V$ is said to **generate** (or **span**) a subspace $W \subseteq V$ if every element of W can be written as a linear combination of $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n$; that is,

$$W \subseteq \text{span}(\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n).$$

Definition 1.4.3. The set of vectors $\{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n\}$ in a vector space V is called a **basis** of V if the vectors are linearly independent and they span V . We say that a vector space V is **finite dimensional** if it has a basis with finitely many elements.

Theorem 1.4.4. Any two bases of a finite-dimensional vector space V have the same number of elements. That is, if $\{\underline{v}_1, \dots, \underline{v}_n\}$ and $\{\underline{w}_1, \dots, \underline{w}_m\}$ are both forming a basis of V , then $n = m$.

The number of vectors in any basis is called the **dimension** of V and denoted by $\dim V$.

Proof. Let

$$\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\} \quad \text{and} \quad \mathcal{C} = \{\underline{w}_1, \dots, \underline{w}_m\}$$

be two bases of the finite-dimensional vector space V . We will prove $n = m$ by showing $n \leq m$ and $m \leq n$.

Since $\mathcal{C}$ is a basis, it spans V . Hence each $\underline{v}_i$ is a linear combination of the $\underline{w}_j$; that is, for every i there exist scalars $a_{1i}, \dots, a_{mi}$ with

$$\underline{v}_i = \sum_{j=1}^m a_{ji} \underline{w}_j. \tag{1.4.1} \text{ ?eq:subthis?}$$

Consider the equation

$$\sum_{i=1}^n x_i \underline{v}_i = \underline{0}.$$

Substituting (1.4.1) in the above equation gives

$$\sum_{i=1}^n x_i \left(\sum_{j=1}^m a_{ji} \underline{w}_j \right) = \sum_{j=1}^m \left(\sum_{i=1}^n a_{ji} x_i \right) \underline{w}_j = \underline{0}.$$

Since the vectors $\underline{w}_1, \dots, \underline{w}_m$ form a basis, each coefficient must vanish:

$$\sum_{i=1}^n a_{ji} x_i = 0 \quad \text{for } j = 1, \dots, m.$$

This is a homogeneous linear system of m equations in the n unknowns $x_1, \dots, x_n$, hence assuming $n > m$, it has a nontrivial solution. Thus if $n > m$ the $\underline{v}_i$ would be linearly dependent,

contradicting that $\mathcal{B}$ is a basis. Therefore $n \leq m$.

The inequality $m \leq n$ can be showed by the above argument with swapping the roles. $\square$

Theorem 1.4.5. Let $\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\}$ be a basis for a vector space V . Then every vector $\underline{v} \in V$ can be written **uniquely** as a linear combination

$$\underline{v} = a_1 \underline{v}_1 + a_2 \underline{v}_2 + \dots + a_n \underline{v}_n,$$

for scalars $a_1, \dots, a_n \in \mathbb{R}$. We use the notation

$$[\underline{v}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

for the representation of $\underline{v}$ in the basis $\mathcal{B}$. These scalars are called the **coordinates** of $\underline{v}$ in the basis $\mathcal{B}$.

Proof. Let us argue by contradiction. Let $\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\}$ be a basis for a vector space V . Since $\mathcal{B}$ spans V , for every $\underline{v} \in V$ there exist scalars $a_1, \dots, a_n \in \mathbb{R}$ such that

$$\underline{v} = a_1 \underline{v}_1 + \dots + a_n \underline{v}_n.$$

Suppose that there exists $\underline{v} \in V$ for which this representation is not unique. That is, there exist scalars $b_1, \dots, b_n \in \mathbb{R}$ with $a_i \neq b_i$ for some $i = 1, \dots, n$ such that

$$\underline{v} = b_1 \underline{v}_1 + \dots + b_n \underline{v}_n.$$

Then using these two linear combinations, we get

$$\underline{0} = \underline{v} - \underline{v} = (a_1 - b_1) \underline{v}_1 + \dots + (a_n - b_n) \underline{v}_n,$$

However, this would mean that there is a non-trivial linear combinations of the vectors $\underline{v}_1, \dots, \underline{v}_n$, which gives the zero vector. This contradicts to the assumption that $\mathcal{B}$ being linearly independent. $\square$

Since every vector in V can be uniquely represented in a basis $\mathcal{B}$, we get

$$[\underline{v} + \underline{w}]_{\mathcal{B}} = [\underline{v}]_{\mathcal{B}} + [\underline{w}]_{\mathcal{B}} \text{ and } [a \cdot \underline{v}]_{\mathcal{B}} = a \cdot [\underline{v}]_{\mathcal{B}}$$

for every $\underline{v}, \underline{w} \in V$ and $a \in \mathbb{R}$.

Example 1.4.6. Consider the vector space of column vectors $\mathbb{R}^n$. The set $\mathcal{N} = \{\underline{e}_1, \dots, \underline{e}_n\}$ forms the natural basis of $\mathbb{R}^n$, where

$$\underline{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \underline{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

In other words, $\underline{e}_k$ is the vector for which the k th element is 1 and every other element is 0.

We will use the convention that if we write $\underline{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$, then the coordinates of $\underline{v}$ are expressed in the natural basis $\mathcal{N}$.

We can rewrite Theorem 1.4.5 in terms of linear equations. The system of linear equations (1.3.1) with coefficients vectors $\underline{v}_1, \dots, \underline{v}_n$ defined in (1.3.2) has a unique solution for every constant vector $\underline{b}$

$$x_1 \underline{v}_1 + \dots + x_n \underline{v}_n = \underline{b}$$

if and only if $\underline{v}_1, \dots, \underline{v}_n$ forms a basis of $\mathbb{R}^n$. So by using Proposition 1.2.7, it follows that $\underline{v}_1, \dots, \underline{v}_n$ forms a basis of $\mathbb{R}^n$ if and only if, after the Gauss-Jordan elimination, the reduced row-echelon form of the matrix

$$A = [\underline{v}_1 \quad \dots \quad \underline{v}_n] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \ddots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

is the $n \times n$ identity matrix

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}.$$

Example 1.4.7. Let $\underline{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3$, Let $\mathcal{B}$ be the basis $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$. What are the coordinates of $\underline{v}$ in basis $\mathcal{B}$?

Let us write for the coordinates of the vector $\underline{v}$ in basis $\mathcal{B}$ by $[\underline{v}]_{\mathcal{B}} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$. Using the definition of coordinates Theorem 1.4.5, we can write $\underline{v}$ as a linear combination of the vectors of the basis. Thus, we have to solve the following linear equation system:

$$x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

We now solve this using Gauss-Jordan elimination.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 3 \end{array} \right] \xrightarrow{R_2 \leftarrow R_2 - R_3} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{array} \right] \xrightarrow{R_1 \leftarrow R_1 - R_3} \left[\begin{array}{ccc|c} 1 & 1 & 0 & -2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{array} \right]$$

$$\xrightarrow{R_1 \leftarrow R_1 - R_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{array} \right].$$

Thus the solution is

$$(x, y, z) = (-1, -1, 3).$$

So the coordinates of $\underline{v}$ in basis $\mathcal{B}$ are:

$$[\underline{v}]_{\mathcal{B}} = \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix}.$$

1.4.1 Finding a basis of the spanned subspace

asis of spanned)?

For a general collection, the vectors $\underline{v}_1, \dots, \underline{v}_n \in \mathbb{R}^m$ usually won't form a basis of $\text{span}\{\underline{v}_1, \dots, \underline{v}_n\}$, because there might be linear dependences. To get rid of the linear dependences is to write it in a matrix form

$$M = \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_n \end{bmatrix}.$$

Perform the Gauss-Jordan elimination on M , we obtain the reduced row-echelon form of M . Suppose that the pivot elements are contained in the columns with indices $j_1, \dots, j_r$. Then

$$\mathcal{B} = \{\underline{v}_{j_1}, \dots, \underline{v}_{j_r}\}$$

is a basis of $\text{span}\{\underline{v}_1, \dots, \underline{v}_n\}$, and in particular, the number of pivot elements is the dimension of the space $\text{span}\{\underline{v}_1, \dots, \underline{v}_n\}$.

To see this, it is enough to show that $\{\underline{v}_{j_1}, \dots, \underline{v}_{j_r}\}$ are linearly independent, and any other vector $\underline{v}_\ell \notin \{\underline{v}_{j_1}, \dots, \underline{v}_{j_r}\}$ can be expressed as a linear combination of $\{\underline{v}_{j_1}, \dots, \underline{v}_{j_r}\}$.

Solve the linear equation

$$x_1 \underline{v}_{j_1} + \cdots + x_r \underline{v}_{j_r} = \underline{0}$$

by repeating the same steps of the Gauss-Jordan elimination performed on M . Hence, we get that every column of the RREF of the matrix

$$\begin{bmatrix} \underline{v}_{j_1} & \cdots & \underline{v}_{j_r} \end{bmatrix}$$

contains a pivot, and so, $x_1 = \cdots = x_r = 0$ is the only solution. This shows the linear independence.

On the other hand, adding any other column vector $\underline{v}_\ell$, where $\ell \neq j_1, \dots, \ell \neq j_r$, and solving the linear equation

$$y_1 \underline{v}_{j_1} + \cdots + y_r \underline{v}_{j_r} + y_{r+1} \underline{v}_\ell = \underline{0}$$

again by repeating the same steps of the Gauss-Jordan elimination performed on M , we see that the all columns of the RREF of

$$\begin{bmatrix} \underline{v}_{j_1} & \cdots & \underline{v}_{j_r} & \underline{v}_\ell \end{bmatrix}$$

contains a pivot element, except the last one. Hence, the equation has infinitely many solutions, and so, the vectors $\{\underline{v}_{j_1}, \dots, \underline{v}_{j_r}, \underline{v}_\ell\}$ are linearly dependent.

As a corollary of the above. The column vectors of the RREF of M which do not contain pivot element are containing the basis representation of the corresponding column in the basis $\mathcal{B}$.

Example 1.4.8. Let

$$\underline{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 3 \end{bmatrix}, \quad \underline{v}_2 = \begin{bmatrix} 2 \\ -5 \\ -3 \\ 6 \end{bmatrix}, \quad \underline{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 0 \end{bmatrix}, \quad \underline{v}_4 = \begin{bmatrix} 2 \\ -1 \\ 4 \\ -7 \end{bmatrix}, \quad \text{and } \underline{v}_5 = \begin{bmatrix} 5 \\ -8 \\ 1 \\ 2 \end{bmatrix}.$$

Find a basis of $U = \text{span}\{\underline{v}_1, \underline{v}_2, \underline{v}_3, \underline{v}_4, \underline{v}_5\}$ out of the vectors $\{\underline{v}_1, \underline{v}_2, \underline{v}_3, \underline{v}_4, \underline{v}_5\}$ and express the remaining vectors in this basis!

We form the 4×5 matrix whose columns are these vectors:

$$A = \begin{bmatrix} 1 & 2 & 0 & 2 & 5 \\ -2 & -5 & 1 & -1 & -8 \\ 0 & -3 & 3 & 4 & 1 \\ 3 & 6 & 0 & -7 & 2 \end{bmatrix}.$$

Apply the Gauss-Jordan elimination to A :

$$\begin{aligned} \begin{bmatrix} 1 & 2 & 0 & 2 & 5 \\ -2 & -5 & 1 & -1 & -8 \\ 0 & -3 & 3 & 4 & 1 \\ 3 & 6 & 0 & -7 & 2 \end{bmatrix} &\xrightarrow{\substack{R_2 \rightarrow R_2 + 2R_1, \\ R_4 \rightarrow R_4 - 3R_1}} \begin{bmatrix} 1 & 2 & 0 & 2 & 5 \\ 0 & -1 & 1 & 3 & 2 \\ 0 & -3 & 3 & 4 & 1 \\ 0 & 0 & 0 & -13 & -13 \end{bmatrix} \\ &\xrightarrow{R_3 \rightarrow R_3 - 3R_2} \begin{bmatrix} 1 & 2 & 0 & 2 & 5 \\ 0 & -1 & 1 & 3 & 2 \\ 0 & 0 & 0 & -5 & -5 \\ 0 & 0 & 0 & -13 & -13 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

From the pivot positions (columns 1, 2, 4), we see that $\mathcal{B} = \{v_1, v_2, v_4\}$ is a basis of U . Thus $\dim U = 3$, and v_3, v_5 can be written as a linear combination of $\underline{v}_1, \underline{v}_2, \underline{v}_4$. Explicitly,

$$[\underline{v}_3]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \quad [\underline{v}_5]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

1.4.2 Equivalent conditions for invertibility

Summarizing of the previous properties, we get the following theorem.

?{thm:inv}?

Theorem 1.4.9 (Equivalent Conditions for Invertibility). Let $A \in \mathbb{R}^{n \times n}$ be an $n \times n$ matrix. The following are equivalent:

1. The reduced row-echelon form of A is the identity matrix.

2. A is invertible.
3. The linear system $A\underline{x} = \underline{0}$ has the unique solution: $\underline{x} = \underline{0}$.
4. The linear system $A\underline{x} = \underline{b}$ has a unique solution for every $\underline{b} \in \mathbb{R}^n$.
5. $\det(A) \neq 0$.
6. The columns of A are linearly independent.
7. The columns of A form a basis of $\mathbb{R}^n$.
8. The rows of A are linearly independent.
9. The rows of A form a basis of $\mathbb{R}^n$.

1.4.3 Change of basis

In a vector space, there are usually many different bases. Next, we discuss how these different bases related to each other.

Let V be a vector space, and let $\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\}$ and $\mathcal{B}' = \{\underline{w}_1, \dots, \underline{w}_n\}$ be two bases of V . Then the **basis transformation matrix** from $\mathcal{B}$ to $\mathcal{B}'$ is the matrix $P_{\mathcal{B}' \rightarrow \mathcal{B}}$ whose columns are the coordinates of $\underline{w}_i$ expressed in the $\mathcal{B}$ basis:

$$P_{\mathcal{B}' \rightarrow \mathcal{B}} := \begin{bmatrix} [\underline{w}_1]_{\mathcal{B}} & [\underline{w}_2]_{\mathcal{B}} & \dots & [\underline{w}_n]_{\mathcal{B}} \end{bmatrix}.$$

Proposition 1.4.10. Let V be a vector space, and let $\mathcal{B}$ and $\mathcal{B}'$ be two bases of V . If a vector $\underline{v} \in V$ has coordinates $[\underline{v}]_{\mathcal{B}'}$ in basis $\mathcal{B}'$, then its coordinates in basis $\mathcal{B}$ are given by the usual matrix multiplication:

$$[\underline{v}]_{\mathcal{B}} = P_{\mathcal{B} \rightarrow \mathcal{B}'} [\underline{v}]_{\mathcal{B}'}.$$

Proof. Let $\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\}$ and $\mathcal{B}' = \{\underline{w}_1, \dots, \underline{w}_n\}$, and let $\underline{v} \in V$ be arbitrary. Let us write the coordinates of $\underline{v}$ in basis $\mathcal{B}'$ as

$$[\underline{v}]_{\mathcal{B}'} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \text{ which means that } b_1 \underline{w}_1 + \dots + b_n \underline{w}_n = \underline{v}.$$

Let us write the coordinates of $\underline{w}_i$ in basis $\mathcal{B}$ by

$$[\underline{w}_i]_{\mathcal{B}} = \begin{bmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{bmatrix} \text{ which means that } a_{1i} \underline{v}_1 + \dots + a_{ni} \underline{v}_n = \underline{w}_i.$$

Hence, combining the above we get

$$\begin{aligned} \underline{v} &= b_1 \underline{w}_1 + \dots + b_n \underline{w}_n \\ &= b_1 (a_{11} \underline{v}_1 + \dots + a_{n1} \underline{v}_n) + \dots + b_n (a_{1n} \underline{v}_1 + \dots + a_{nn} \underline{v}_n) \\ &= (a_{11} b_1 + \dots + a_{1n} b_n) \underline{v}_1 + \dots + (a_{n1} b_1 + \dots + a_{nn} b_n) \underline{v}_n. \end{aligned}$$

Using the definition of the coordinates of the vector $\underline{v}$ in basis $\mathcal{B}$, we get

$$[\underline{v}]_{\mathcal{B}} = \begin{bmatrix} a_{11}b_1 + \cdots + a_{1n}b_n \\ \vdots \\ a_{n1}b_1 + \cdots + a_{nn}b_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix},$$

which had to be proven. $\square$

A simple consequence of the uniqueness of the coordinate representation of vectors Proposition 1.4.10 is the following: let $\mathcal{B}$, $\mathcal{B}'$ and $\mathcal{B}''$ be bases of a vector space V . Then the basis transformation matrices satisfy the following identity:

$$P_{\mathcal{B} \rightarrow \mathcal{B}''} = P_{\mathcal{B}' \rightarrow \mathcal{B}''} P_{\mathcal{B} \rightarrow \mathcal{B}'}.$$

In particular, for any two bases $\mathcal{B}$, $\mathcal{B}'$ of V

$$I = P_{\mathcal{B} \rightarrow \mathcal{B}'} P_{\mathcal{B}' \rightarrow \mathcal{B}},$$

where I is the identity matrix. Thus, $P_{\mathcal{B} \rightarrow \mathcal{B}'} = (P_{\mathcal{B}' \rightarrow \mathcal{B}})^{-1}$ with the usual matrix inversion.

Example 1.4.11. Let $V = \mathbb{R}^2$. Let $\mathcal{N} = \left\{ \underline{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \underline{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ be the natural basis, and let $\mathcal{B} = \left\{ \underline{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \underline{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ be another basis.

(a) Find the basis transformation matrices $P_{\mathcal{N} \rightarrow \mathcal{B}}$ and $P_{\mathcal{B} \rightarrow \mathcal{N}}$.

(b) Let $\underline{v} = [\underline{v}]_{\mathcal{N}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. Find the coordinates $[\underline{v}]_{\mathcal{B}}$ of the vector $\underline{v}$ in basis $\mathcal{B}$.

We wish to find the change of basis matrix $P_{\mathcal{B} \rightarrow \mathcal{N}}$ from $\mathcal{B}$ to $\mathcal{N}$. Since $\underline{v}_1$ and $\underline{v}_2$ are expressed in the natural basis, we have $P_{\mathcal{B} \rightarrow \mathcal{N}}$ as the matrix of the column vectors of the base $\mathcal{B}$ (in the natural basis), i.e.

$$P_{\mathcal{B} \rightarrow \mathcal{N}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

On the other hand, we have

$$P_{\mathcal{N} \rightarrow \mathcal{B}} = (P_{\mathcal{B} \rightarrow \mathcal{N}})^{-1} = \frac{1}{-2} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

and so

$$[\underline{v}]_{\mathcal{B}} = P_{\mathcal{N} \rightarrow \mathcal{B}} [\underline{v}]_{\mathcal{N}} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

1.4.4 Exercises

1. Let us consider the following vectors: $\underline{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\underline{u}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$ and $\underline{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix}$. Are the vectors $\underline{u}_1, \underline{u}_2, \underline{u}_3$ linearly independent? If not express the one with the others!
2. Let $\mathcal{B} = \left\{ \underline{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \underline{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^2$, and let $\underline{v} = \begin{bmatrix} -1 \\ 5 \end{bmatrix}$. Find the coordinates of $\underline{v}$ in the basis $\mathcal{B}$.
3. Let $\mathcal{B}$ be the basis of $\mathbb{R}^2$ as in the previous exercise. Find the basis transformation matrices $P_{\mathcal{B} \rightarrow \mathcal{N}}$ and $P_{\mathcal{N} \rightarrow \mathcal{B}}$, where $\mathcal{N}$ is the natural basis of $\mathbb{R}^2$.

1.5 Linear transformations

Definition 1.5.1. Let V and W be vector spaces over $\mathbb{R}$. A function $T : V \rightarrow W$ is called a **linear transformation** (or a **linear map**) if, for all $\underline{u}, \underline{v} \in V$ and $\lambda \in \mathbb{R}$, the following properties hold:

1. **Additivity:** $T(\underline{u} + \underline{v}) = T(\underline{u}) + T(\underline{v})$,
2. **Homogeneity:** $T(\lambda \underline{v}) = \lambda T(\underline{v})$.

A simple consequence that a linear map T maps the $\underline{0}$ vector (of the vectorspace V) to the zero vector $\underline{0}$ (of the vector space W). Indeed,

$$T(\underline{0}) = T(\underline{0} + \underline{0}) = T(\underline{0}) + T(\underline{0}) = 2T(\underline{0}).$$

Thus, $T(\underline{0}) = \underline{0}$.

Example 1.5.2. The following maps are linear transformations on the plane $\mathbb{R}^2$:

- Reflection across the x -axis:

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix}.$$

- Reflection across the line $y = x$:

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

- Projection onto the x -axis:

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix}$$

Non-example: The transformation $T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x+1 \\ y \end{bmatrix}$ is **not** linear, since it does not preserve the zero vector: $T \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

For every $m \times n$ matrix A , the action $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ defined as

$$T(\underline{v}) = A\underline{v}$$

gives a linear transformation.

Definition 1.5.3 (Matrix Representation of a Linear Transformation). *Let $T: V \rightarrow W$ be a linear transformation between finite-dimensional vector spaces. Let $\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\}$ be a basis of V , and let $\mathcal{C} = \{\underline{w}_1, \dots, \underline{w}_m\}$ be a basis of W . Then the matrix of T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$, denoted $[T]_{\mathcal{B} \rightarrow \mathcal{C}}$, is the $m \times n$ matrix whose j -th column is the coordinate vector of $T(\underline{v}_j)$ with respect to $\mathcal{C}$, that is:*

$$[T]_{\mathcal{B} \rightarrow \mathcal{C}} = \begin{bmatrix} [T(\underline{v}_1)]_{\mathcal{C}} & \cdots & [T(\underline{v}_n)]_{\mathcal{C}} \end{bmatrix}.$$

If $V = W$ and $\mathcal{B}$ is a basis of V then we use the simplified notation $[T]_{\mathcal{B}} = [T]_{\mathcal{B} \rightarrow \mathcal{B}}$

The matrix representation of the linear map T satisfies:

$$[T(\underline{v})]_{\mathcal{C}} = [T]_{\mathcal{B} \rightarrow \mathcal{C}} \cdot [\underline{v}]_{\mathcal{B}} \quad \text{for all } \underline{v} \in V. \quad (1.5.1) \text{ ?eq:linrep?}$$

To see this, let us write $\underline{v} \in V$ in the basis $\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\}$, i.e. $[\underline{v}]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$. Then

$$T(\underline{v}) = T(b_1\underline{v}_1 + \cdots + b_n\underline{v}_n) = b_1T(\underline{v}_1) + \cdots + b_nT(\underline{v}_n).$$

Thus,

$$[T(\underline{v})]_{\mathcal{C}} = [b_1T(\underline{v}_1) + \cdots + b_nT(\underline{v}_n)]_{\mathcal{C}} = b_1[T(\underline{v}_1)]_{\mathcal{C}} + \cdots + b_n[T(\underline{v}_n)]_{\mathcal{C}},$$

which implies (1.5.1).

Example 1.5.4. Let $T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ be the transformation defined by:

$$T(x, y, z) = (2x - y, x + z).$$

Find the matrix representation $[T]_{\mathcal{N}_3 \rightarrow \mathcal{N}_2}$ of T in the natural bases of $\mathbb{R}^2$ and $\mathbb{R}^3$.

Let $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ be the natural basis of $\mathbb{R}^3$ and $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ the natural basis of $\mathbb{R}^2$. Then

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \text{and} \quad T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

So the matrix representation of T is:

$$[T]_{\mathcal{N}_3 \rightarrow \mathcal{N}_2} = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Theorem 1.5.5 (Change of Basis for Linear Transformations). *Let $T : V \rightarrow V$ be a linear transformation, and let $\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\}$ and $\mathcal{B}' = \{\underline{w}_1, \dots, \underline{w}_n\}$ be two bases of V . Then*

$$[T]_{\mathcal{B}'} = P_{\mathcal{B} \rightarrow \mathcal{B}'} [T]_{\mathcal{B}} P_{\mathcal{B}' \rightarrow \mathcal{B}}.$$

Proof. Using the definition of the basis representation of the linear transformation T in (1.5.1) and the property of the basis transformation matrix Proposition 1.4.10, we have

$$[T]_{\mathcal{B}'} [\underline{x}]_{\mathcal{B}'} = [T(\underline{x})]_{\mathcal{B}'} = P_{\mathcal{B} \rightarrow \mathcal{B}'} [T(\underline{x})]_{\mathcal{B}} = P_{\mathcal{B} \rightarrow \mathcal{B}'} [T]_{\mathcal{B}} [\underline{x}]_{\mathcal{B}} = P_{\mathcal{B} \rightarrow \mathcal{B}'} [T]_{\mathcal{B}} P_{\mathcal{B}' \rightarrow \mathcal{B}} [\underline{x}]_{\mathcal{B}'}.$$

The matrices on the left-hand side and the right-hand side are both satisfying (1.5.1), thus, are equal. $\square$

Example 1.5.6. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^2$, and let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be a linear transformation defined by:

$$T(x, y) = (3x + y, x + 2y).$$

What is the matrix representation $[T]_{\mathcal{B}}$ of T in basis $\mathcal{B}$?

In the standard basis, the matrix of T is:

$$[T]_{\mathcal{N}} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}.$$

To compute the matrix of T in the new basis $\mathcal{B}$ we construct the basis transformation matrices $P_{\mathcal{N} \rightarrow \mathcal{B}}, P_{\mathcal{B} \rightarrow \mathcal{N}}$:

$$P_{\mathcal{B} \rightarrow \mathcal{N}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ and } P_{\mathcal{N} \rightarrow \mathcal{B}} = (P_{\mathcal{B} \rightarrow \mathcal{N}})^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

Compute the new matrix representation:

$$[T]_{\mathcal{B}} = (P_{\mathcal{B} \rightarrow \mathcal{N}})^{-1} [T]_{\mathcal{N}} P_{\mathcal{B} \rightarrow \mathcal{N}} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}.$$

Let us now make a slight detour and show that the determinant of a square matrix $A \in \mathbb{R}^{n \times n}$ has a nice geometric interpretation.

?<rem:det>?

Remark 1.5.7 (Geometric Meaning of the Determinant). *Let A be an $n \times n$ matrix of reals. Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a linear transformation defined as $T(\underline{x}) = A\underline{x}$. For a region $U \subset \mathbb{R}^n$, let*

$$T(U) = \{T(\underline{x}) : \underline{x} \in U\}$$

be the image of the region U under the linear transformation T . Then

$$\text{Vol}(T(U)) = |\det(A)| \text{Vol}(U),$$

where Vol denotes the volume (area in $\mathbb{R}^2$). That is, $|\det(A)|$ tells how the transformation T scales the volume of any region in $\mathbb{R}^n$.

The **sign of the determinant** indicates whether the transformation **preserves orientation**:

- $\det(A) > 0$: orientation is preserved,
- $\det(A) < 0$: orientation is reversed,
- $\det(A) = 0$: the transformation collapses volume entirely (i.e., maps n -dimensional volume into a lower-dimensional space).

1.5.1 Exercises

1. Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the linear transformation that rotates the plane with 60 deg around the origin in the counter clockwise direction. What is the basis representation matrix of T in the natural basis?
2. Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the linear transformation defined as

$$T(x, y) = (2x - y, 3y - x).$$

Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^2$. What is the basis representation of T in the natural basis and in basis $\mathcal{B}$?

1.6 Eigenvalues, eigenvectors

Now we wish to find a basis, where the linear transformation acts nicely.

Definition 1.6.1. Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. A non-zero vector $\underline{v} \in \mathbb{R}^n \setminus \{\underline{0}\}$ is called an **eigenvector** of A if there exists a scalar $\lambda \in \mathbb{R}$ such that

$$A\underline{v} = \lambda\underline{v}.$$

The scalar λ is called the **eigenvalue** corresponding to the **eigenvector** $\underline{v}$.

The following theorem provides a way of finding the eigenvalues of an $n \times n$ matrix A .

Theorem 1.6.2. Let A be an $n \times n$ matrix. Then λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$. This is called the **characteristic equation** of A .

Proof. If λ is an eigenvalue of A then there exists a non-zero vector $\underline{v} \in \mathbb{R}^n$ such that $A\underline{v} = \lambda\underline{v}$. Then

$$A\underline{v} = \lambda\underline{v} \Rightarrow A\underline{v} - \lambda\underline{v} = \underline{0} \Rightarrow (A - \lambda I)\underline{v}.$$

Hence, the matrix $A - \lambda I$ is not invertible and by Corollary 1.1.15, $\det(A - \lambda I) = 0$.

On the other hand, if $\det(A - \lambda I) = 0$ then by Theorem 1.4.9 $A - \lambda I$ is not invertible and there exists $\underline{v}$ non-zero vector $\underline{v} \in \mathbb{R}^n \setminus \{\underline{0}\}$ such that $(A - \lambda I)\underline{v} = \underline{0}$, and so, $A\underline{v} = \lambda\underline{v}$, which means that λ is an eigenvalue. $\square$

Example 1.6.3. Consider the matrix $A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$. Find the eigenvalues and eigenvectors of A .

We compute the eigenvalues by solving the characteristic equation:

$$\det \left(\begin{bmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{bmatrix} \right) = (4-\lambda)(3-\lambda) - 2 = \lambda^2 - 7\lambda + 10 = 0.$$

The solutions are $\lambda_1 = 5$, $\lambda_2 = 2$. To find the eigenvector for $\lambda_1 = 5$, we need to solve the linear equation:

$$(A - \lambda_1 I)\underline{v} = \underline{0} \quad \Leftrightarrow \quad \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad x = y \quad \Rightarrow \quad \underline{v}_1 = \begin{bmatrix} x \\ y \end{bmatrix}.$$

For $\lambda_2 = 2$, solve:

$$\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad y = -2x \quad \Rightarrow \quad \underline{v}_2 = \begin{bmatrix} x \\ -2x \end{bmatrix}.$$

As the example shows, an eigenvalue has many eigenvectors. In particular, the eigenvectors of A corresponding to the eigenvalue λ forms a subspace.

Definition 1.6.4. We call an $n \times n$ matrix $A \in \mathbb{R}^{n \times n}$ **diagonalizable** if there exists eigenvectors $\underline{v}_1, \dots, \underline{v}_n$ of A which form a basis of $\mathbb{R}^n$.

Example 1.6.5. Consider the matrix

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

Its characteristic polynomial is

$$\det(A - \lambda I) = (\lambda - 1)^2,$$

so the only eigenvalue is $\lambda = 1$. To find eigenvectors, we solve

$$(A - I)\underline{v} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} = \underline{0}.$$

This gives $y = 0$, so every eigenvector is of the form $\underline{v} = \begin{bmatrix} x \\ 0 \end{bmatrix}$. Hence, the eigenvectors cannot form a basis of $\mathbb{R}^2$, and for diagonalization we would need a basis, thus A cannot be diagonalized.

?<thm:diag>?

Theorem 1.6.6 (Eigendecomposition). Let $A \in \mathbb{R}^{n \times n}$ be diagonalizable. Then there exists an invertible matrix P and a diagonal matrix D such that

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

where

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \text{ and } P = P_{\mathcal{B} \rightarrow \mathcal{N}} = \begin{bmatrix} \underline{v}_1 & \cdots & \underline{v}_n \end{bmatrix},$$

where $\lambda_1, \dots, \lambda_n$ are eigenvalues of A , and $\underline{v}_i$ is the eigenvector corresponding to the eigenvalue λ_i for $i = 1, \dots, n$.

Proof. Since A is diagonalizable, there exists a basis $\mathcal{B} = \{\underline{v}_1, \dots, \underline{v}_n\}$ of $\mathbb{R}^n$ consisting of eigenvectors of A . By definition of an eigenvector, we have

$$A\underline{v}_i = \lambda_i \underline{v}_i, \quad i = 1, \dots, n,$$

where λ_i is the eigenvalue corresponding to $\underline{v}_i$.

Now, let P be the basis transformation matrix

$$P := P_{\mathcal{B} \rightarrow \mathcal{N}} = \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_n \end{bmatrix}.$$

Next, consider the action of AP :

$$\begin{aligned} AP &= A \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_n \end{bmatrix} = \begin{bmatrix} A\underline{v}_1 & A\underline{v}_2 & \cdots & A\underline{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 \underline{v}_1 & \lambda_2 \underline{v}_2 & \cdots & \lambda_n \underline{v}_n \end{bmatrix} = \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} = PD, \end{aligned}$$

where $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$ is the diagonal matrix formed by the eigenvalues. Multiplying

both sides by P^{-1} on the right gives $A = PDP^{-1}$. $\square$

This means in particular that the matrix representation of the diagonalisable matrix A in the basis formed by the eigenvectors of A is a diagonal matrix.

Corollary 1.6.7. *If $A \in \mathbb{R}^{n \times n}$ has eigenvalues $\lambda_1, \dots, \lambda_n$, then*

$$\det(A) = \prod_{i=1}^n \lambda_i.$$

Proof. We show it only in the case when A is diagonalisable. Then by Theorem 1.6.6 and the properties of the determinant Theorem 1.1.10

$$\det(A) = \det(PDP^{-1}) = \det(P) \det(D) \det(P^{-1}) = \det(D),$$

hence, the claim follows. $\square$

1.6.1 Exercises

1. Diagonalise the matrix $A = \begin{bmatrix} 14 & -4 \\ 30 & -9 \end{bmatrix}$.
2. Find the eigenvalues of the matrix $A = \begin{bmatrix} 2 & -1 & -1 \\ 0 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix}$.

1.7 Scalar product

Now, we equip the vector space V with the so-called scalar product. This allows us to study the geometric properties of vectors, like length or angle.

Definition 1.7.1. Let V be a vector space over $\mathbb{R}$. A function

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$$

is called a **scalar product** on V if it satisfies the following properties for all $\underline{u}, \underline{v}, \underline{w} \in V$ and all scalars $\lambda \in \mathbb{R}$:

1. **Symmetry:** $\langle \underline{u}, \underline{v} \rangle = \langle \underline{v}, \underline{u} \rangle$.
2. **Linearity in the first argument:**

$$\langle \underline{u} + \underline{v}, \underline{w} \rangle = \langle \underline{u}, \underline{w} \rangle + \langle \underline{v}, \underline{w} \rangle \text{ and } \langle \lambda \underline{u}, \underline{w} \rangle = \lambda \langle \underline{u}, \underline{w} \rangle.$$

3. **Positive definiteness:**

$$\langle \underline{v}, \underline{v} \rangle \geq 0 \quad \text{and} \quad \langle \underline{v}, \underline{v} \rangle = 0 \text{ if and only if } \underline{v} = \underline{0}.$$

The pair $(V, \langle \cdot, \cdot \rangle)$ is called an **inner product space**.

Example 1.7.2 (Scalar Product on $\mathbb{R}^n$). Let $\underline{u} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \underline{v} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$. Their **scalar product** is

$$\langle \underline{u}, \underline{v} \rangle = \underline{u}^T \underline{v} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{i=1}^n x_i y_i.$$

Definition 1.7.3.

- The **length** (or **norm**) of a vector $\underline{v} \in V$ is defined by

$$\|\underline{v}\| = \sqrt{\langle \underline{v}, \underline{v} \rangle}.$$

- The **angle** $\angle(\underline{u}, \underline{v})$ between non-zero vectors $\underline{u}$ and $\underline{v}$ is given by

$$\angle(\underline{u}, \underline{v}) = \arccos \left(\frac{\langle \underline{u}, \underline{v} \rangle}{\|\underline{u}\| \cdot \|\underline{v}\|} \right).$$

- We say that the vectors $\underline{u}, \underline{v} \in V$ are called **orthogonal** if

$$\langle \underline{u}, \underline{v} \rangle = 0.$$

Definition 1.7.4. A basis $\{\underline{v}_1, \dots, \underline{v}_n\} \subset \mathbb{R}^n$ is called an **orthonormal basis** if its elements are pairwise orthogonal and have length one. That is,

$$\langle \underline{v}_i, \underline{v}_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Definition 1.7.5. A square matrix $Q \in \mathbb{R}^{n \times n}$ is called **orthogonal** if $Q^{-1} = Q^\top$, or equivalently, $Q^\top Q = I$.

Theorem 1.7.6. Let $Q \in \mathbb{R}^{n \times n}$ be an orthogonal matrix. Then

1. The columns of Q form an orthonormal basis of $\mathbb{R}^n$.
2. Q preserves scalar products: for all $u, v \in \mathbb{R}^n$,

$$\langle Qu, Qv \rangle = \langle u, v \rangle.$$

3. Q preserves lengths and angles.

4. $|\det Q| = 1$.

Proof. Suppose that $Q = [\underline{q}_1 \ \cdots \ \underline{q}_n]$ is an orthogonal matrix, where $\underline{q}_i$ are the column vectors.

Then $Q^\top = \begin{bmatrix} \underline{q}_1^\top \\ \vdots \\ \underline{q}_n^\top \end{bmatrix}$ and so

$$Q^\top Q = \begin{bmatrix} \underline{q}_1^\top \\ \vdots \\ \underline{q}_n^\top \end{bmatrix} [\underline{q}_1 \ \cdots \ \underline{q}_n] = \begin{bmatrix} \underline{q}_1^\top \underline{q}_1 & \cdots & \underline{q}_1^\top \underline{q}_n \\ \vdots & \ddots & \vdots \\ \underline{q}_n^\top \underline{q}_1 & \cdots & \underline{q}_n^\top \underline{q}_n \end{bmatrix}.$$

Hence, $\underline{q}_i^\top \underline{q}_j = 1$ if $i = j$ and $\underline{q}_i^\top \underline{q}_j = 0$ if $i \neq j$.

On the other hand, for every $\underline{u}, \underline{v} \in \mathbb{R}^n$

$$\langle Qu, Qv \rangle = (Qu)^\top Qv = \underline{u}^\top Q^\top Qv = \underline{u}^\top \underline{v},$$

hence, Q preserves the length and so the angles.

Finally,

$$1 = \det(I) = \det(QQ^\top) = (\det Q)^2.$$

□

1.8 Symmetric matrices

Definition 1.8.1. We call an $n \times n$ matrix A **symmetric**, if $A^\top = A$.

(prop:symdiag)?

Proposition 1.8.2. Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then all eigenvalues of A are real. Moreover, for every $\underline{v}_1, \underline{v}_2 \in \mathbb{R}^n$ eigenvectors of A corresponding to distinct eigenvalues $\lambda_1 \neq \lambda_2$, $\underline{v}_1$ and $\underline{v}_2$ are orthogonal.

Proof. We only show the second claim. Assume $A\underline{v}_1 = \lambda_1\underline{v}_1$ and $A\underline{v}_2 = \lambda_2\underline{v}_2$, with $\lambda_1 \neq \lambda_2$. Then

$$\lambda_1 \langle \underline{v}_1, \underline{v}_2 \rangle = \langle \lambda_1 \underline{v}_1, \underline{v}_2 \rangle = \langle A\underline{v}_1, \underline{v}_2 \rangle = \langle \underline{v}_1, A^\top \underline{v}_2 \rangle = \langle \underline{v}_1, A\underline{v}_2 \rangle = \langle \underline{v}_1, \lambda_2 \underline{v}_2 \rangle = \lambda_2 \langle \underline{v}_1, \underline{v}_2 \rangle.$$

Hence

$$(\lambda_1 - \lambda_2) \langle \underline{v}_1, \underline{v}_2 \rangle = 0.$$

Since $\lambda_1 \neq \lambda_2$, it follows that $\langle \underline{v}_1, \underline{v}_2 \rangle = 0$. □

(thm:symdiag)?

Theorem 1.8.3. Every real symmetric matrix is orthogonally diagonalizable. That is, if $A \in \mathbb{R}^{n \times n}$ and $A^\top = A$, then there exists an orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ and a diagonal matrix $D \in \mathbb{R}^{n \times n}$ such that

$$A = QDQ^\top.$$

The columns of Q are eigenvectors of A , which form an orthonormal basis, and the diagonal entries of D are the corresponding eigenvalues.

Example 1.8.4. Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. Diagonalise the matrix!

Let us first find the roots of the characteristic polynomial, which are the eigenvalues.

$$0 = \det(A - \lambda I) = (2 - \lambda)^2 - 1 \cdot 1 = \lambda^2 - 4\lambda + 3,$$

therefore the eigenvalues are

$$\lambda_{1,2} = \frac{4 \pm \sqrt{4^2 - 4 \cdot 3}}{2} = \begin{cases} 3, \\ 1. \end{cases}$$

We obtain the eigenvectors through solving the equation

$$(A - 3I)\underline{x} = \underline{0}, \quad \text{that is,} \quad \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underline{0}.$$

This reduces to $x_1 - x_2 = 0$, i.e. $x_1 = x_2$. So an eigenvector is

$$\underline{v}_1 = \begin{bmatrix} x \\ x \end{bmatrix}.$$

By normalizing it, we get

$$1 = \|\underline{v}_1\|^2 = x^2 + x^2 \Rightarrow x = \pm \frac{\sqrt{2}}{2},$$

and so $\underline{v}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$. We know that the eigenvectors of a symmetric matrix A are pairwise orthogonal, then on the plane, we can choose the eigenvector $\underline{v}_2$ of the eigenvalue $\lambda_2 = 1$ as

$$\underline{v}_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}.$$

So,

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

Theorem 1.8.3 follows essentially from Proposition 1.8.2, however, it might happen that the eigenvalues of A appears with multiplicity. That is, the eigenvectors of a certain eigenvalue forms a subspace with dimension strictly greater than two. To handle this case, we use the Gram-Schmidt orthogonalisation on that subspace.

1.8.1 Gram-Schmidt orthogonalisation

Algorithm 1.8.5 (Gram–Schmidt Orthogonalization). *Let V be a vector space. Let $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_n$ be linearly independent vectors. The Gram–Schmidt process constructs an orthonormal set of vectors $\underline{u}_1, \underline{u}_2, \dots, \underline{u}_n$ that spans the same subspace. First, let us define vectors $\underline{u}'_1, \dots, \underline{u}'_n$ recursively. Let*

$$\underline{u}'_1 := \underline{v}_1.$$

Then, let

$$\underline{u}_2 := \underline{v}_2 - \frac{\langle \underline{v}_2, \underline{u}'_1 \rangle}{\langle \underline{u}'_1, \underline{u}'_1 \rangle} \underline{u}'_1.$$

If the vectors $\underline{u}'_1, \dots, \underline{u}'_k$ are defined for $k < n$ then let

$$\underline{u}'_{k+1} := \underline{v}_{k+1} - \frac{\langle \underline{v}_{k+1}, \underline{u}'_1 \rangle}{\langle \underline{u}'_1, \underline{u}'_1 \rangle} \underline{u}'_1 - \dots - \frac{\langle \underline{v}_{k+1}, \underline{u}'_k \rangle}{\langle \underline{u}'_k, \underline{u}'_k \rangle} \underline{u}'_k.$$

Finally, for every $k = 1, \dots, n$ let

$$\underline{u}_k := \frac{\underline{u}'_k}{\|\underline{u}'_k\|}.$$

The algorithm indeed produces an orthonormal basis. The vectors $\underline{u}_1, \underline{u}_2, \dots, \underline{u}_n$ clearly have unit length. Since the angle does not depend on the length, it is enough to check that the vectors $\underline{u}'_1, \underline{u}'_2, \dots, \underline{u}'_n$ are pairwise orthogonal. We show the orthogonality inductively. Suppose that the vectors $\underline{u}'_1, \underline{u}'_2, \dots, \underline{u}'_k$ are pairwise orthogonal for some $k < n$. Then for every $i < k + 1$

$$\begin{aligned} \langle \underline{u}'_i, \underline{u}'_{k+1} \rangle &= \left\langle \underline{u}'_i, \underline{v}_{k+1} - \frac{\langle \underline{v}_{k+1}, \underline{u}'_1 \rangle}{\langle \underline{u}'_1, \underline{u}'_1 \rangle} \underline{u}'_1 - \dots - \frac{\langle \underline{v}_{k+1}, \underline{u}'_k \rangle}{\langle \underline{u}'_k, \underline{u}'_k \rangle} \underline{u}'_k \right\rangle \\ &= \langle \underline{u}'_i, \underline{v}_{k+1} \rangle - \frac{\langle \underline{v}_{k+1}, \underline{u}'_1 \rangle}{\langle \underline{u}'_1, \underline{u}'_1 \rangle} \langle \underline{u}'_i, \underline{u}'_1 \rangle - \dots - \frac{\langle \underline{v}_{k+1}, \underline{u}'_k \rangle}{\langle \underline{u}'_k, \underline{u}'_k \rangle} \langle \underline{u}'_i, \underline{u}'_k \rangle \\ &= \langle \underline{u}'_i, \underline{v}_{k+1} \rangle - \frac{\langle \underline{v}_{k+1}, \underline{u}'_i \rangle}{\langle \underline{u}'_i, \underline{u}'_i \rangle} \langle \underline{u}'_i, \underline{u}'_i \rangle = 0, \end{aligned}$$

where at the equation before the last, we used the inductive assumption on pairwise orthogonality.

Example 1.8.6. Find an orthonormal basis of the vectorspace $V = \text{span}\{\underline{v}_1, \underline{v}_2, \underline{v}_3\}$, where

$$\underline{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \underline{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \underline{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Let us apply the Gram-Schmidt orthogonalisation on the vectors $\{\underline{v}_1, \underline{v}_2, \underline{v}_3\}$, which form a basis of V . In the first step, let $\underline{u}'_1 := \underline{v}_1$. So,

$$\|\underline{u}'_1\| = \sqrt{1^2 + 1^2 + 0^2 + 0^2} \text{ and } \underline{u}_1 = \frac{\underline{v}_1}{\|\underline{v}_1\|} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix}.$$

For the second basis vector, let:

$$\begin{aligned} \underline{u}'_2 &:= \underline{v}_2 - \frac{\langle \underline{v}_2, \underline{u}'_1 \rangle}{\langle \underline{u}'_1, \underline{u}'_1 \rangle} \underline{u}'_1 = \underline{v}_2 - \langle \underline{v}_2, \underline{u}_1 \rangle \underline{u}_1 \\ &= \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \left(1 \cdot \frac{1}{\sqrt{2}} + 0 \cdot \frac{1}{\sqrt{2}} + 1 \cdot 0 + 0 \cdot 0 \right) \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \\ 0 \end{bmatrix}. \end{aligned}$$

The length of $\underline{u}'_2$ is $\|\underline{u}'_2\| = \sqrt{(\frac{1}{2})^2 + (-\frac{1}{2})^2 + 1^2 + 0^2} = \sqrt{\frac{3}{2}}$, so by normalising $\underline{u}'_2$, we get that the second element of the basis is

$$\underline{u}_2 = \frac{\underline{u}'_2}{\|\underline{u}'_2\|} = \frac{1}{\sqrt{\frac{3}{2}}} \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ 0 \end{bmatrix}.$$

Finally, we compute the third basis vector. By definition,

$$\begin{aligned} \langle \underline{v}_3, \underline{u}_1 \rangle &= 1 \cdot \frac{1}{\sqrt{2}} + 0 \cdot \frac{1}{\sqrt{2}} + 0 \cdot 0 + 1 \cdot 0 = \frac{1}{\sqrt{2}}, \\ \langle \underline{v}_3, \underline{u}_2 \rangle &= 1 \cdot \frac{1}{\sqrt{6}} + 0 \cdot \left(-\frac{1}{\sqrt{6}} \right) + 0 \cdot \frac{2}{\sqrt{6}} + 1 \cdot 0 = \frac{1}{\sqrt{6}}. \end{aligned}$$

Let

$$\underline{u}'_3 = \underline{v}_3 - \langle \underline{v}_3, \underline{u}_1 \rangle \underline{u}_1 - \langle \underline{v}_3, \underline{u}_2 \rangle \underline{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix} - \frac{1}{\sqrt{6}} \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 - \frac{1}{2} - \frac{1}{6} \\ 0 - \frac{1}{2} + \frac{1}{6} \\ 0 - 0 - \frac{2}{6} \\ 1 - 0 - 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \\ 1 \end{bmatrix}.$$

Now,

$$\|\underline{u}'_3\| = \sqrt{\left(\frac{1}{3}\right)^2 + \left(-\frac{1}{3}\right)^2 + \left(-\frac{1}{3}\right)^2 + 1^2} = \frac{2}{\sqrt{3}}$$

and so,

$$\underline{u}_3 = \frac{\underline{u}'_3}{\|\underline{u}'_3\|} = \frac{\sqrt{3}}{2} \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{6} \\ -\frac{\sqrt{3}}{6} \\ -\frac{\sqrt{3}}{6} \\ \frac{\sqrt{3}}{2} \end{bmatrix}.$$

So the orthonormal basis is:

$$\left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\sqrt{3}}{6} \\ -\frac{\sqrt{3}}{6} \\ -\frac{\sqrt{3}}{6} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \right\}.$$

1.8.2 Quadratic Forms

Now, we will consider an application of symmetric matrices.

Definition 1.8.7. A **quadratic form** of two variables is a polynomial with terms all of degree 2. That is,

$$Q(x, y) = ax^2 + bxy + cy^2,$$

where a, b, c are real numbers.

A quadratic form can be written as

$$Q(x, y) = ax^2 + bxy + cy^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

We intend to study the level sets of quadratic forms. That is, we wish to understand and draw the set of points (x, y) on the plane such that

$$Q(x, y) = d$$

for some $d \in \mathbb{R}$. These are called conic sections (ellipses, hyperbolas, pairs of lines, etc.). The

kind of the shape depends on the signs of the eigenvalues of the matrix

$$A = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}.$$

To see this, we apply Theorem 1.8.3 and diagonalise the symmetric matrix A . That is, $A = QDQ^\top$, where $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ and $Q = [\underline{u}_1 \quad \underline{u}_2]$ are such that $A\underline{u}_i = \lambda_i\underline{u}_i$ and $\{\underline{u}_1, \underline{u}_2\}$ forms an orthonormal basis of $\mathbb{R}^2$. Then

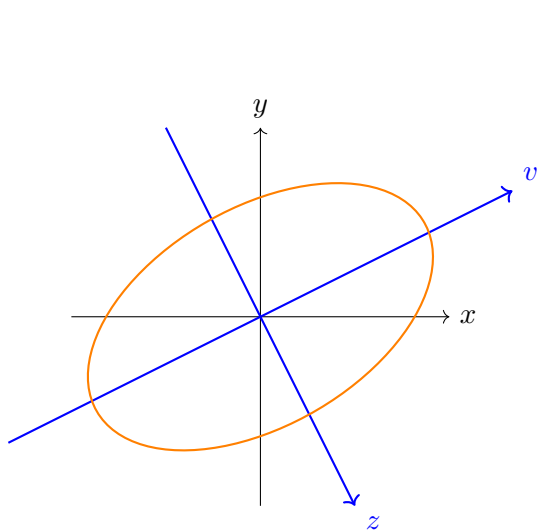
$$Q(x, y) = \begin{bmatrix} x & y \end{bmatrix} QDQ^\top \begin{bmatrix} x \\ y \end{bmatrix} = \left(P^\top \begin{bmatrix} x \\ y \end{bmatrix} \right)^\top D \left(P^\top \begin{bmatrix} x \\ y \end{bmatrix} \right).$$

Let us write the vector $\begin{bmatrix} x \\ y \end{bmatrix}$ in the new coordinate system formed by $\{\underline{u}_1, \underline{u}_2\}$ as $\begin{bmatrix} v \\ z \end{bmatrix}$, i.e. $P^\top \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} v \\ z \end{bmatrix}$. Hence, the level set is

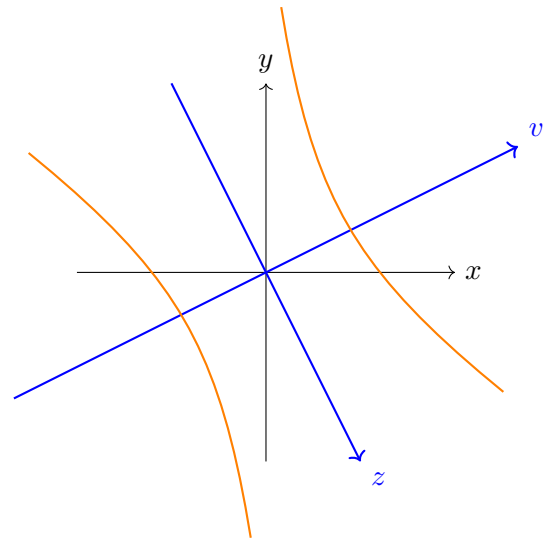
$$d = Q(v, z) = \begin{bmatrix} v & z \end{bmatrix} D \begin{bmatrix} v \\ z \end{bmatrix} = \lambda_1 v^2 + \lambda_2 z^2$$

in the coordinate system formed by $\{\underline{u}_1, \underline{u}_2\}$. From this form, we see that

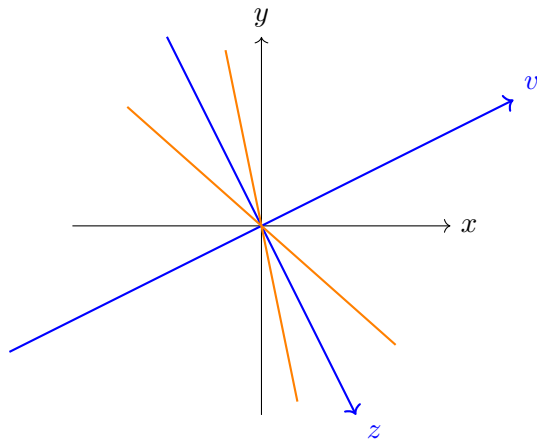
1. If $\lambda_1 \lambda_2 > 0$ and $\lambda_1 d > 0$ then it forms an ellipse which intersects the z axis at $\pm \sqrt{\frac{d}{\lambda_2}}$ and the v axis at $\pm \sqrt{\frac{d}{\lambda_1}}$;
2. If $\lambda_1 \lambda_2 > 0$ and $\lambda_1 d = 0$ then it is one point, the origin, and if $\lambda_1 d < 0$ then it is the empty set;
3. If $\lambda_1 = 0$ and $\lambda_2 d > 0$ then it is two lines parallel to the v axis intersecting the z axis at $\pm \sqrt{\frac{d}{\lambda_2}}$;
4. If $\lambda_1 = 0$ and $d = 0$ then it is the v axis, and if $d \lambda_2 < 0$ then it is the empty set;
5. If $\lambda_2 = 0$ and $\lambda_1 d > 0$ then it is two lines parallel to the z axis intersecting the v axis at $\pm \sqrt{\frac{d}{\lambda_1}}$;
6. If $\lambda_2 = 0$ and $d = 0$ then it is the z axis, and if $d \lambda_1 < 0$ then it is the empty set;
7. If $\lambda_1 \lambda_2 < 0$ and $\lambda_1 d > 0$ then it is a hyperbola intersecting the v axis at $\pm \sqrt{\frac{d}{\lambda_1}}$;
8. If $\lambda_1 \lambda_2 < 0$ and $\lambda_2 d > 0$ then it is a hyperbola intersecting the z axis at $\pm \sqrt{\frac{d}{\lambda_2}}$;
9. If $\lambda_1 \lambda_2 < 0$ and $d = 0$ then it two lines intersecting at the origin slopes $\pm \frac{\lambda_1}{\lambda_2}$ with respect to the v axis.



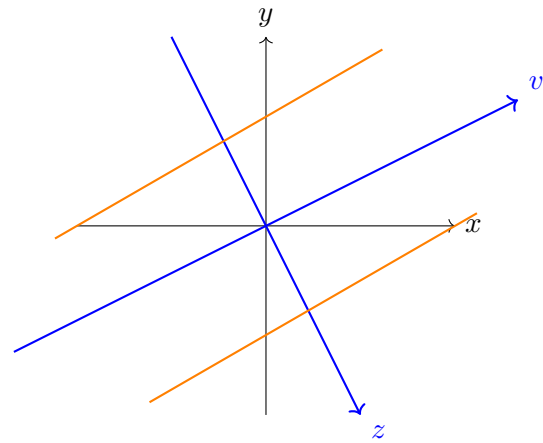
(a) The ellipse in case $\lambda_1 \lambda_2 > 0$ and $\lambda_1 d > 0$ intersecting the v axis at $\pm\sqrt{\frac{d}{\lambda_1}}$ and the z axis at $\pm\sqrt{\frac{d}{\lambda_2}}$.



(b) The hyperbola in case $\lambda_1 \lambda_2 < 0$ and $\lambda_1 d > 0$ intersecting the v axis at $\pm\sqrt{\frac{d}{\lambda_1}}$.



(c) The two lines in case $\lambda_1 \lambda_2 < 0$ and $d = 0$ with slopes $\frac{\lambda_1}{\lambda_2}$ with in point of view of v axis.



(d) The two lines in case $\lambda_1 = 0$ and $d\lambda_2 > 0$ intersecting the z axis at $\pm\sqrt{\frac{d}{\lambda_2}}$.

Figure 1.2: The possible cases of the quadratic form.

Example 1.8.8. Draw a sketch of the points (x, y) on the plane which satisfy the equation

$$2x^2 + 2y^2 - 2xy = 1.$$

First, we write this quadratic form as a matrix product:

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

For matrix $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$, we solve the characteristic equation to find the eigenvalues.

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

Therefore: $\lambda_1 = 1$ and $\lambda_2 = 3$. For $\lambda_1 = 1$:

$$(A - I)\underline{v}_1 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This gives us $x - y = 0$, so $x = y$, and so $\underline{v}_1 = \begin{bmatrix} y \\ y \end{bmatrix}$. For normalising the length, we need $y^2 + y^2 = 1$,

thus, $\underline{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$. Since the eigenvectors are pairwise orthogonal, we get that the eigenvector for

$\lambda_2 = 3$ can be chosen as $\underline{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$. Hence, in the coordinate system $\{\underline{v}_1, \underline{v}_2\}$ our equation becomes

$$u^2 + 3v^2 = 1$$

This is the equation of an ellipse in standard form with semi-axes of length $a = 1$ (along the u -axis) and $b = \frac{1}{\sqrt{3}}$ (along the v -axis).

Conclusion: The original equation describes an ellipse rotated 45 counterclockwise from the coordinate axes, with major axis length 1 along direction $(1, 1)$ and minor axis length $\frac{1}{\sqrt{3}}$ along direction $(1, -1)$.

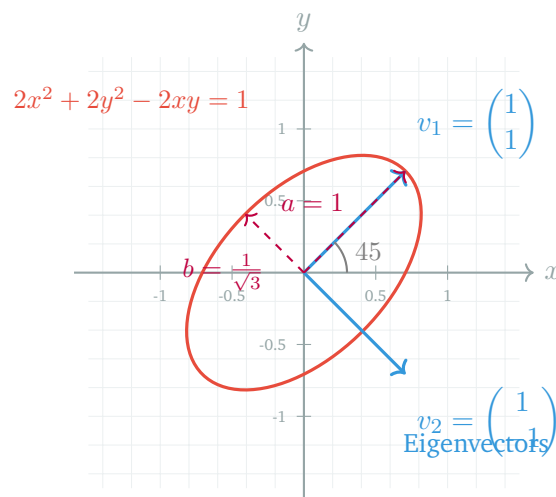


Figure 1.3: Visualization of the quadratic form $2x^2 + 2y^2 - 2xy = 1$. The red ellipse is rotated 45 from the coordinate axes. The blue vectors show the eigenvectors, which align with the principal axes of the ellipse. The purple dashed lines indicate the major axis ($a = 1$) and minor axis ($b = 1/\sqrt{3}$) in the rotated coordinate system.

1.8.3 Exercises

1. Diagonalise the matrix $A = \begin{bmatrix} 7 & -2 \\ -2 & 3 \end{bmatrix}$.
2. Using Gram-Schmidt orthogonalisation, find an orthonormal basis of the subspace

$$\left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 : x - 3y + z = 0 \right\}.$$

3. Draw a sketch of the points on the plane which satisfies the equation $9x^2 + 4xy + 6y^2 = 5$.
4. Diagonalise the 3×3 matrix $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}$. (Hint: Use Gram-Schmidt orthogonalisation for the eigenvalue which has multiplicity 2.)

1.9 Trace and double dot product*

Definition 1.9.1. The **trace** of a square matrix $A \in \mathbb{R}^{n \times n}$, denoted by $\text{tr}(A)$, is defined as the sum of the elements on the main diagonal:

$$\text{tr}(A) = \sum_{i=1}^n a_{ii}.$$

Theorem 1.9.2. Basic Properties of the Trace Let $A, B, C \in \mathbb{R}^{n \times n}$ be square matrices, and let $\alpha \in \mathbb{R}$. Then:

- $\text{tr}(A^T) = \text{tr}(A)$
- $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$,
- $\text{tr}(\alpha A) = \alpha \cdot \text{tr}(A)$,
- $\text{tr}(AB) = \text{tr}(BA)$,
- More generally, for three matrices:

$$\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB),$$

but **not necessarily** equal to $\text{tr}(ACB)$ or any non-cyclic permutation.

Theorem 1.9.3. Let $A \in \mathbb{R}^{n \times n}$ be a diagonalisable matrix with eigenvalues $\lambda_1, \dots, \lambda_n$. Then:

$$\text{tr}(A) = \sum_{i=1}^n \lambda_i.$$

Proof. Since A is diagonalisable, by Theorem 1.6.6 there exist an invertible matrix P and a diagonal

matrix D such that $A = PDP^{-1}$ and $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$. Then by the last assertion of

Theorem 1.9.2,

$$\operatorname{tr}(A) = \operatorname{tr}(PDP^{-1}) = \operatorname{tr}(P^{-1}PD) = \operatorname{tr}(D) = \sum_{i=1}^n \lambda_i.$$

□

Definition 1.9.4. Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \ddots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \ddots & b_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$ be two real $m \times n$ matrices. The **double-dot product** is defined as

$$A : B := \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ij}.$$

Equivalently,

$$A : B = \operatorname{tr}(A^T B).$$

Theorem 1.9.5. Let $A, B, C \in \mathbb{R}^{m \times n}$ and $\alpha \in \mathbb{R}$. Then:

1. **Symmetry:**

$$A : B = B : A.$$

2. **Bilinearity:**

$$(\alpha A + B) : C = \alpha(A : C) + (B : C),$$

$$A : (\alpha B + C) = \alpha(A : B) + (A : C).$$

3. **Positivity:**

$$A : A = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 \geq 0,$$

with equality if and only if $A = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$.

Definition 1.9.6. An $n \times n$ matrix $K \in \mathbb{R}^{n \times n}$ is called **skew-symmetric** (or **antisymmetric**) if

$$K^T = -K.$$

Equivalently, $k_{ii} = 0$ for all i and $k_{ij} = -k_{ji}$ for all $i \neq j$.

Theorem 1.9.7. For every square matrix $A \in \mathbb{R}^{n \times n}$ there exist unique matrices S and K such that

$$A = S + K,$$

where S is symmetric ($S^\top = S$) and K is skew-symmetric ($K^\top = -K$). Moreover,

$$S = \frac{A + A^\top}{2}, \quad K = \frac{A - A^\top}{2}.$$

1.9.1 Exercises

1.9/1. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$. Find $\text{tr}(A) = ?$ and $A : B = ?$

1.9/2. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$. Write A as $A = S + K$, where S is a symmetric and K is a skew-symmetric matrix.

1.10 Fundamental Subspaces of a Matrix

Definition 1.10.1. Let $A \in \mathbb{R}^{m \times n}$. The **column space** of A is the subspace of $\mathbb{R}^m$ spanned by its columns:

$$\text{Col}(A) = \text{span}\{\underline{c}_1, \underline{c}_2, \dots, \underline{c}_n\},$$

where $\underline{c}_j \in \mathbb{R}^m$ denotes the j -th column vector of A .

Definition 1.10.2. Let $A \in \mathbb{R}^{m \times n}$. The **row space** of A is the subspace of $\mathbb{R}^n$ spanned by its rows:

$$\text{Row}(A) = \text{span}\{\underline{r}_1^\top, \underline{r}_2^\top, \dots, \underline{r}_m^\top\},$$

where $\underline{r}_j \in \mathbb{R}^n$ denotes the j -th row vector of A .

Observe that $\text{Row}(A) = \text{Col}(A^\top)$.

Theorem 1.10.3. For any matrix $A \in \mathbb{R}^{m \times n}$,

$$\dim \text{Col}(A) = \dim \text{Row}(A).$$

This common dimension is called the **rank** of A , and denoted by $\text{rank}(A)$.

Proof. Apply Gauss-Jordan elimination to bring A into reduced row echelon form. It is clear that the elementary row manipulations does not change the row space. Moreover, the non-zero rows will be linearly independent due to the strict ordering between the pivot elements (see third claim in the Definition 1.2.4). So, the non-zero rows of the RREF of A will form a basis of $\text{Row}(A)$. In particular, the number of pivot elements equals $\dim \text{Row}(A)$.

By Section 1.4.1, the columns that contain a pivot element in the RREF of A , form a basis of $\text{Col}(A)$, and so, the number of pivot elements equals $\dim \text{Col}(A)$. $\square$

Definition 1.10.4. The **nullspace** (or **kernel**) of the $m \times n$ matrix A is the subspace

$$\text{Null}(A) = \{\underline{x} \in \mathbb{R}^n \mid A\underline{x} = \underline{0}\}.$$

We denote the dimension of $\text{Null}(A)$ by $\text{nullity}(A) = \dim \text{Null}(A)$.

Definition 1.10.5. For an $m \times n$ matrix $A \in \mathbb{R}^{m \times n}$, the subspaces $\text{Col}(A)$, $\text{Row}(A)$, $\text{Null}(A)$ and $\text{Null}(A^\top)$ the **fundamental subspaces** of the matrix A .

$\langle \text{thm:ranknull} \rangle?$

Theorem 1.10.6 (Rank–nullity theorem). Let $A \in \mathbb{R}^{m \times n}$ be an $m \times n$ real matrix. Then:

$$\text{rank}(A) + \text{nullity}(A) = n.$$

Proof. Let us solve the linear equation $A\underline{x} = \underline{0}$, where $\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Apply the Gauss-Jordan elimination, and let M be the reduced row-echelon form of A . Suppose that the columns $j_1, \dots, j_k$ contain a pivot element, and the columns $i_1, \dots, i_{n-k}$ does not contain pivot. Hence, the solution of the equation $A\underline{x} = \underline{0}$ can be written in the form

$$x_{j_1} = - \sum_{\ell=1}^{n-k} m_{j_1, i_\ell} x_{i_\ell}, \dots, x_{j_k} = - \sum_{\ell=1}^{n-k} m_{j_k, i_\ell} x_{i_\ell}, \quad (1.10.1) \text{ ?eq:nullsp?}$$

where $m_{j,i}$ denotes the element of the RREF M in the j th row and i th column.

Hence, the vectors $\underline{x}_1, \dots, \underline{x}_{n-k}$ form a basis of $\text{Null}(A)$ where the elements of $\underline{x}_p$ are such that $x_{i_p} = 1$, $x_{i_\ell} = 0$ if $\ell \neq p$ and the remaining elements satisfy (1.10.1). Since the number of pivot elements corresponds to $\text{rank}(A) = k$, the claim follows. $\square$

Theorem 1.10.7. For all $A \in \mathbb{R}^{m \times n}$, we have

$$\text{rank}(A) = \text{rank}(A^\top A).$$

Proof. Let A be an $m \times n$ real matrix. Observe that $A^\top A$ is an $n \times n$ matrix. It is enough to show that $\text{Null}(A) = \text{Null}(A^\top A)$. Then the claim follows by Theorem 1.10.6 applied for both A and $A^\top A$.

If $\underline{x} \in \text{Null}(A)$ then $A\underline{x} = \underline{0}$, and so, $A^\top A\underline{x} = \underline{0}$ and so, $\underline{x} \in \text{Null}(A^\top A)$. On the other hand, if $\underline{x} \in \text{Null}(A^\top A)$ then $A^\top A\underline{x} = \underline{0}$. Then

$$0 = \underline{x}^\top A^\top A\underline{x} = (A\underline{x})^\top A\underline{x} = \|A\underline{x}\|^2.$$

This implies that $A\underline{x} = \underline{0}$, which gives that $\underline{x} \in \text{Null}(A)$. $\square$

Example 1.10.8. Let

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

- (a) Give a basis of the column space $\text{Col}(A)$ out of the column vectors of A , and express the coordinates of the remaining vectors in this basis.
- (b) Find a basis of the null space $\text{Null}(A)$.
- (c) What are $\text{rank}(A)$, $\text{rank}(A^\top A)$, $\text{nullity}(A)$ and $\text{nullity}(A^\top)$?

First, let's find the reduced row echelon form (RREF) of A by using the Gauss-Jordan elimination. Now, we omit the details and leave the calculation for the reader.:

We get

$$\begin{bmatrix} 0 & 2 & 4 & 0 & 2 \\ 1 & 1 & 1 & 3 & 2 \\ 2 & -1 & -4 & 0 & -5 \\ -1 & 0 & 1 & 2 & 4 \end{bmatrix} \xrightarrow{\text{by Gauss-Jordan elimination}} \begin{bmatrix} 1 & 0 & -1 & 0 & -2 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

- (a) The columns that contain a pivot element are the columns 1, 2, and 4. Therefore, a basis of $\text{Col}(A)$ is:

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

From the RREF form, we can express the remaining columns:

$$\begin{bmatrix} 4 \\ 1 \\ -4 \\ 1 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 2 \\ 2 \\ -5 \\ 4 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}.$$

- (b) For $\text{Null}(A)$, we solve $Ax = \mathbf{0}$. From the RREF, we see that:

$$\begin{cases} x_1 - x_3 - 2x_5 = 0 \\ x_2 + 2x_3 + x_5 = 0 \\ x_4 + x_5 = 0 \end{cases}$$

Setting free variables $x_3 = s$ and $x_5 = t$:

$$\begin{cases} x_1 = s + 2t \\ x_2 = -2s - t \\ x_3 = s \\ x_4 = -t \\ x_5 = t \end{cases}$$

Therefore every vector $\underline{v} \in \text{Null}(A)$ can be written as

$$\underline{x} = s \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

Thus, a basis of $\text{Null}(A)$ is:

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

(c) From our calculations:

- $\text{rank}(A) = 3$ (number of pivot columns)
- $\text{rank}(A^\top A) = \text{rank}(A) = 3$
- $\text{nullity}(A) = 5 - 3 = 2$ (by rank-nullity theorem)
- $\text{nullity}(A^\top) = 4 - 3 = 1$ (since $\text{rank}(A^\top) = \text{rank}(A) = 3$ and again by the rank-nullity theorem)

1.10.1 Orthogonal complements

Definition 1.10.9. Let V be a vector space and $W \subseteq V$ a subspace. The **orthogonal complement** of W is

$$W^\perp := \{\underline{v} \in V : \underline{v} \text{ is perpendicular to } \underline{w} \text{ for all } \underline{w} \in W\}.$$

<thm:perpfund>?

Theorem 1.10.10 (Fundamental subspaces and orthogonal complements). For any $A \in \mathbb{R}^{m \times n}$:

$$\text{Row}(A)^\perp = \text{Null}(A), \quad \text{Null}(A^\top) = \text{Col}(A)^\perp.$$

Proof. Let $A = \begin{bmatrix} \underline{r}_1^\top \\ \vdots \\ \underline{r}_m^\top \end{bmatrix}$ be an $m \times n$ real matrix where $\underline{r}_i$ denotes the i th row vector. If $\underline{x} \in \text{Null}(A)$ then for every $i = 1, \dots, m$, $\underline{r}_i^\top \underline{x} = 0$, and in particular, $\underline{x}$ is perpendicular for every row vector of A . This implies that $\underline{x}$ is perpendicular for every linear combination of the row vectors, and so, $\underline{x} \in \text{Row}(A)^\perp$. The other direction is straightforward, since if $\underline{x} \in \text{Row}(A)^\perp$ then for every $i = 1, \dots, m$, $\underline{r}_i^\top \underline{x} = 0$, and so, $A\underline{x} = \underline{0}$.

For the second statement, consider $A^\top$. □

<thm:ortocom>?

Theorem 1.10.11 (Dimension and orthogonal complements). Let $W \subseteq \mathbb{R}^n$ be a subspace. Then

$$\dim(W) + \dim(W^\perp) = n.$$

Proof. Choose a basis $\{\underline{w}_1, \dots, \underline{w}_k\}$ of W , and let $A = \begin{bmatrix} \underline{w}_1^\top \\ \vdots \\ \underline{w}_k^\top \end{bmatrix}$ be an $k \times n$ matrix formed by these vectors as column vectors. Hence, $\text{Row}(A) = W$, and in particular, $\text{rank}(A) = \dim W$.

By Theorem 1.10.10, $\text{Null}(A) = \text{Row}(A)^\perp = W^\perp$. Thus, $\text{nullity}(A) = \dim W^\perp$. The claim then follows by Theorem 1.10.6. $\square$

Example 1.10.12. Let

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

Determine all four fundamental subspaces and check the statement of the Theorem 1.10.10 on them.

For that we will use Gauss-Jordan elimination. Starting with A :

$$\begin{bmatrix} 1 & -3 & -2 \\ 2 & -6 & -4 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 - 2R_1} \begin{bmatrix} 1 & -3 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

This is already in reduced row echelon form.

- Column space $\text{Col}(A)$: The pivot column of A is column 1. Therefore:

$$\text{Col}(A) = \text{span} \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

- Row space $\text{Row}(A)$: The nonzero row of $\text{RREF}(A)$ gives us:

$$\text{Row}(A) = \text{span} \left\{ \begin{bmatrix} 1 \\ -3 \\ -2 \end{bmatrix} \right\}$$

- Null space $\text{Null}(A)$: From $\text{RREF}(A)$, we solve $A\mathbf{x} = \mathbf{0}$:

$$x_1 - 3x_2 - 2x_3 = 0$$

Setting free variables $x_2 = s$ and $x_3 = t$:

$$\begin{cases} x_1 = 3s + 2t \\ x_2 = s \\ x_3 = t \end{cases}$$

Therefore:

$$\underline{x} = s \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$

So:

$$\text{Null}(A) = \text{span} \left\{ \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Finally, to determine $\text{Null}(A^\top)$, we need to apply the Gauss-Jordan elimination on $A^\top$.

$$A^\top = \begin{bmatrix} 1 & 2 \\ -3 & -6 \\ -2 & -4 \end{bmatrix} \xrightarrow{\substack{R_2 \leftarrow R_2 + 3R_1 \\ R_3 \leftarrow R_3 + 2R_1}} \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus,

- The null space $\text{Null}(A^\top)$: The solutions of $A^\top \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{0}$ are

$$y_1 + 2y_2 = 0.$$

Setting free variable $y_2 = s$:

$$\begin{cases} y_1 = -2s \\ y_2 = s \end{cases}$$

Therefore:

$$\underline{y} = s \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

So:

$$\text{Null}(A^\top) = \text{span} \left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}$$

Let us now verify that $\text{Row}(A)^\perp = \text{Null}(A)$. We need to show that every vector in $\text{Null}(A)$ is orthogonal to every vector in $\text{Row}(A)$. For that, it is enough to check for the orthogonality of the base vectors.

Row space basis vector: $\underline{r} = \begin{bmatrix} 1 \\ -3 \\ -2 \end{bmatrix}$. Null space basis vectors: $\underline{n}_1 = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$, $\underline{n}_2 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$. Then

$$\langle \underline{r}, \underline{n}_1 \rangle = (1)(3) + (-3)(1) + (-2)(0) = 3 - 3 + 0 = 0 \checkmark$$

$$\langle \underline{r}, \underline{n}_2 \rangle = (1)(2) + (-3)(0) + (-2)(1) = 2 + 0 - 2 = 0 \checkmark$$

Furthermore, $\dim(\text{Row}(A)) = 1$ and $\dim(\text{Null}(A)) = 2$, and so, by Theorem 1.10.11, we see that $\text{Null}(A)^\perp = \text{Row}(A)$.

Finally, we verify that $\text{Null}(A^\top) = \text{Col}(A)^\perp$. To show that every vector in $\text{Null}(A^\top)$ is orthogonal to every vector in $\text{Col}(A)$, again, it is enough to check for the orthogonality of the base vectors.

Column space basis vector: $\underline{c} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Left null space basis vector: $\underline{l} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$. Then

$$\langle \underline{c}, \underline{l} \rangle = (1)(-2) + (2)(1) = -2 + 2 = 0 \checkmark$$

Also, $\dim(\text{Col}(A)) = 1$ and $\dim(\text{Null}(A^\top)) = 1$, so $1 + 1 = 2 = m$, confirming that $\text{Null}(A^\top) = \text{Col}(A)^\perp$.

1.10.2 Exercises

1.10/1. Let

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

- (a) Give a basis of the column space $\text{Col}(A)$ out of the column vectors of A , and express the coordinates of the remaining vectors in this basis.
- (b) What are $\text{rank}(A)$, $\text{rank}(A^\top A)$, $\text{nullity}(A)$ and $\text{nullity}(A^\top)$?

1.10/2. Let A be the matrix as in the previous exercise.

- (a) Find a basis of $\text{Null}(A)$.
- (b) Find an orthonormal basis of $\text{Null}(A)$.

1.11 Orthogonal projections

Now, we will study a special class of linear transformations, namely, the orthogonal projections.

Definition 1.11.1. Let V be a finite dimensional vector space with a scalar product. Furthermore, let $W \subseteq V$ be a subspace, and let $\underline{v} \in V$ be a vector. Then the **orthogonal projection** of $\underline{v}$ onto W is the unique vector $\underline{w} \in W$ such that $\underline{v} - \underline{w} \in W^\perp$. We denote the map, which maps $\underline{v}$ to its the orthogonal projection, by P_W .

Note that the orthogonal projection of a vector $\underline{v}$ is well defined, since if there would be $\underline{w}_1, \underline{w}_2 \in W$ such that $\underline{v} - \underline{w}_1 \in W^\perp$ and $\underline{v} - \underline{w}_2 \in W^\perp$. Then using that $W^\perp$ is a subspace, we get $(\underline{v} - \underline{w}_1) - (\underline{v} - \underline{w}_2) = \underline{w}_2 - \underline{w}_1 \in W^\perp$. But $\underline{w}_2 - \underline{w}_1 \in W$, which means that $\underline{w}_1 = \underline{w}_2$.

An important property about the orthogonal projections is that for a subspace $W \subseteq \mathbb{R}^n$, the closest vector to $\underline{v} \in \mathbb{R}^n$ in W is its orthogonal projection $P_W(\underline{v})$.

?<thm:projmin>?

Theorem 1.11.2. Let $W \subseteq \mathbb{R}^n$ be a subspace and let $P_W: \mathbb{R}^n \rightarrow W$ be the orthogonal projection. Then for every $\underline{v} \in \mathbb{R}^n$

$$\min_{\underline{w} \in W} \|\underline{v} - \underline{w}\| = \|\underline{v} - P_W(\underline{v})\|.$$

Proof. Let $\underline{v} \in \mathbb{R}^n$ and $\underline{w} \in W$ be arbitrary. Then

$$\begin{aligned}\|\underline{v} - \underline{w}\|^2 &= \langle \underline{v} - \underline{w}, \underline{v} - \underline{w} \rangle = \langle \underline{v} - P_W(\underline{v}) + P_W(\underline{v}) - \underline{w}, \underline{v} - P_W(\underline{v}) + P_W(\underline{v}) - \underline{w} \rangle \\ &= \langle \underline{v} - P_W(\underline{v}), \underline{v} - P_W(\underline{v}) \rangle + 2\langle \underline{v} - P_W(\underline{v}), P_W(\underline{v}) - \underline{w} \rangle + \langle P_W(\underline{v}) - \underline{w}, P_W(\underline{v}) - \underline{w} \rangle.\end{aligned}$$

Since $P_W(\underline{v}) - \underline{w} \in W$ and $\underline{v} - P_W(\underline{v}) \in W^\perp$, we get that this is equal to

$$= \|\underline{v} - P_W(\underline{v})\|^2 + \|P_W(\underline{v}) - \underline{w}\|^2 \geq \|\underline{v} - P_W(\underline{v})\|^2.$$

Thus, the minimum of $\|\underline{v} - \underline{w}\|^2$ is attained precisely at $\underline{w} = P_W(\underline{v})$. $\square$

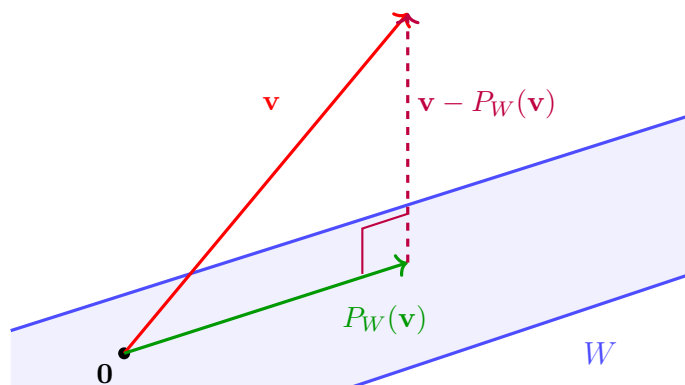


Figure 1.4: Visualisation of the orthogonal projection from $\mathbb{R}^3$ to the plane W .

First, let us study the special case of orthogonal projections to one-dimensional subspaces.

op:planarproj)?

Proposition 1.11.3. Let $\underline{u} \in \mathbb{R}^n$ be a non-zero vector, and let $U = \text{span}\{\underline{u}\}$. The orthogonal projection of a vector $\underline{v} \in \mathbb{R}^n$ onto U (or the vector $\underline{u}$) is the vector

$$P_U(\underline{v}) = \frac{\langle \underline{u}, \underline{v} \rangle}{\langle \underline{u}, \underline{u} \rangle} \underline{u} = \frac{\underline{u} \underline{u}^\top}{\underline{u}^\top \underline{u}} \underline{v}.$$

Proof. Write $\underline{v} = \alpha \underline{u} + \underline{w}$, where $\underline{w}$ is perpendicular to $\underline{u}$. Then the orthogonality condition becomes:

$$0 = \langle \underline{w}, \underline{u} \rangle = \langle \underline{v} - \alpha \underline{u}, \underline{u} \rangle = \langle \underline{v}, \underline{u} \rangle - \alpha \langle \underline{u}, \underline{u} \rangle.$$

Hence, choosing $\alpha = \frac{\langle \underline{u}, \underline{v} \rangle}{\langle \underline{u}, \underline{u} \rangle}$, gives the desired formula. $\square$

Example 1.11.4. Let $\underline{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\underline{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. Find the orthogonal projection of $\underline{v}$ to the subspace spanned by $\underline{u}$.

Applying Proposition 1.11.3, we get that the orthogonal projection is

$$\frac{\langle \underline{u}, \underline{v} \rangle}{\langle \underline{u}, \underline{u} \rangle} \underline{u} = \frac{2 \cdot 3 + 1 \cdot 4}{2^2 + 1^2} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \frac{10}{5} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$

Observe that for a given non-zero vector $\underline{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$, the matrix

$$P = \frac{\underline{u} \underline{u}^\top}{\underline{u}^\top \underline{u}} = \frac{1}{u_1^2 + \dots + u_n^2} \begin{bmatrix} u_1 u_1 & \dots & u_1 u_n \\ \vdots & \ddots & \vdots \\ u_n u_1 & \dots & u_n u_n \end{bmatrix}$$

is the matrix representation of the orthogonal projection $\underline{v} \mapsto P_{\text{span}\{\underline{u}\}} \underline{v}$. In the following, we construct the matrix representation for general subspace $W \subseteq \mathbb{R}^n$.

p:ortprojform)?

Proposition 1.11.5. Let $W \subseteq \mathbb{R}^n$ be a subspace. Let $\underline{u}_1, \underline{u}_2, \dots, \underline{u}_k \in \mathbb{R}^n$ be a basis of W and write

$$M = \begin{bmatrix} \underline{u}_1 & \dots & \underline{u}_k \end{bmatrix}$$

the matrix formed by the basis $\underline{u}_1, \dots, \underline{u}_k$ as column vectors. Then the orthogonal projection P_W is given by the matrix

$$P_W = M(M^\top M)^{-1} M^\top.$$

We call the matrix P_W the **orthogonal projection matrix** (in the natural basis).

Proof. By definition, $\text{Col}(M) = W$, and so, $W^\perp = \text{Col}(M)^\perp = \text{Null}(M^\top)$ by Theorem 1.10.10. Furthermore, every $\underline{w} \in W$ can be expressed uniquely as a linear combination of vectors $\{\underline{u}_1, \dots, \underline{u}_k\}$. In particular, for every $\underline{v} \in \mathbb{R}^n$ there exists $\underline{c} \in \mathbb{R}^k$ such that $P_W(\underline{v}) = M\underline{c}$. Since $\underline{v} - P_W(\underline{v}) = \underline{v} - M\underline{c} \in W^\perp$, we get

$$M^\top(\underline{v} - M\underline{c}) = \underline{0}.$$

This gives that $M^\top M \underline{c} = M^\top \underline{v}$. Since $M^\top M$ is a $k \times k$ matrix with $\text{rank}(M^\top M) = \text{rank}(M) = k$, we get that it is invertible by Theorem 1.4.9, and so

$$\underline{c} = (M^\top M)^{-1} M^\top \underline{v}.$$

Thus,

$$P_W(\underline{v}) = M\underline{c} = M(M^\top M)^{-1} M^\top \underline{v}.$$

□

Example 1.11.6. Find the matrix of the orthogonal projection from $\mathbb{R}^3$ to the plane $V = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : 2x - y + 3z = 0 \right\}$.

What is the orthogonal projection of $\underline{v} = \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix}$ onto V ?

We choose two base vectors in V . For instance,

$$\underline{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \underline{u}_2 = \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix},$$

since $2 \cdot 1 - 2 + 3 \cdot 0 = 0$ and $2(-3) - 0 + 3 \cdot 2 = -6 + 6 = 0$. Form the matrix $M \in \mathbb{R}^{3 \times 2}$ whose

columns are these basis vectors:

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

Compute

$$(M^\top M)^{-1} = \frac{1}{56} \begin{bmatrix} 13 & 3 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 13/56 & 3/56 \\ 3/56 & 5/56 \end{bmatrix}.$$

Hence the matrix of the orthogonal projection onto V is

$$P_V = M(M^\top M)^{-1}M^\top = \begin{bmatrix} \frac{5}{7} & \frac{1}{7} & -\frac{3}{7} \\ \frac{1}{7} & \frac{13}{14} & \frac{3}{14} \\ -\frac{3}{7} & \frac{3}{14} & \frac{5}{14} \end{bmatrix}.$$

Now project the vector $\underline{v} = \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix}$ to V as:

$$P_V \underline{v} = \begin{bmatrix} \frac{5}{7} & \frac{1}{7} & -\frac{3}{7} \\ \frac{1}{7} & \frac{13}{14} & \frac{3}{14} \\ -\frac{3}{7} & \frac{3}{14} & \frac{5}{14} \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{17}{7} \\ \frac{53}{14} \\ -\frac{5}{14} \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 34 \\ 53 \\ -5 \end{bmatrix}.$$

Let us give here an alternative solution too: Observe that plane V is the orthogonal complement of its normal vector

$$\underline{n} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix},$$

hence and by Proposition 1.11.3, the orthogonal projection matrix onto V is

$$P_V = I - \frac{\underline{n}\underline{n}^\top}{\|\underline{n}\|^2}.$$

We have $\|\underline{n}\|^2 = 2^2 + (-1)^2 + 3^2 = 14$ and

$$\underline{n}\underline{n}^\top = \begin{bmatrix} 4 & -2 & 6 \\ -2 & 1 & -3 \\ 6 & -3 & 9 \end{bmatrix}.$$

Hence

$$P_V = I - \frac{1}{14} \begin{bmatrix} 4 & -2 & 6 \\ -2 & 1 & -3 \\ 6 & -3 & 9 \end{bmatrix} = \begin{bmatrix} 1 - \frac{4}{14} & -\frac{-2}{14} & -\frac{6}{14} \\ -\frac{-2}{14} & 1 - \frac{1}{14} & -\frac{-3}{14} \\ -\frac{6}{14} & -\frac{-3}{14} & 1 - \frac{9}{14} \end{bmatrix} = \begin{bmatrix} \frac{5}{7} & \frac{1}{7} & -\frac{3}{7} \\ \frac{1}{7} & \frac{13}{14} & \frac{3}{14} \\ -\frac{3}{7} & \frac{3}{14} & \frac{5}{14} \end{bmatrix}.$$

Therefore

$$P_V(\underline{v}) = P_V \cdot \underline{v} = \begin{bmatrix} \frac{5}{7} & \frac{1}{7} & -\frac{3}{7} \\ \frac{1}{7} & \frac{13}{14} & \frac{3}{14} \\ -\frac{3}{7} & \frac{3}{14} & \frac{5}{14} \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 34 \\ 53 \\ -5 \end{bmatrix}$$

The following theorem characterises when a matrix is an orthogonal projection.

Theorem 1.11.7. *Let $P \in \mathbb{R}^{n \times n}$ be an $n \times n$ matrix. Then P is the orthogonal projection matrix onto the subspace $\text{Col}(P)$ if and only if*

$$P^2 = P \quad \text{and} \quad P^\top = P.$$

Proof. First, suppose that P is an orthogonal projection onto a subspace $W = \text{Col}(P)$. Then using the form $P = M(M^\top M)^{-1}M^\top$ in Proposition 1.11.5, one can see that $P^2 = P$ and $P^\top = P$.

Now suppose that P satisfies the properties $P^2 = P$ and $P^\top = P$. To show that P is a matrix of an orthogonal projection, it is enough to show that $\underline{v} - P\underline{v}$ is perpendicular to $P\underline{v}$. Indeed,

$$\langle \underline{v} - P\underline{v}, P\underline{v} \rangle = \langle \underline{v}, P\underline{v} \rangle - \langle P\underline{v}, P\underline{v} \rangle = \langle \underline{v}, P\underline{v} \rangle - \langle \underline{v}, P^\top P\underline{v} \rangle = \langle \underline{v}, P\underline{v} \rangle - \langle \underline{v}, P^2 \underline{v} \rangle = \langle \underline{v}, P\underline{v} \rangle - \langle \underline{v}, P\underline{v} \rangle = 0,$$

which had to be shown. $\square$

1.11.1 Method of least squares

We have seen that a linear equation $A\underline{x} = \underline{b}$ has no solution if and only if $\underline{b} \notin \text{Col}(A)$. In this case, we can study the vector $\underline{x}$, which is closest to be a solution in the sense that $\|A\underline{x} - \underline{b}\|$ is minimal.

Theorem 1.11.8 (Approximation via projections). *Let $A \in \mathbb{R}^{m \times n}$ be an $m \times n$ matrix and let $\underline{b} \in \mathbb{R}^m$. If the linear equation system $A\underline{x} = \underline{b}$ has no solution, then the minimizer $\underline{x}^*$ of $\|A\underline{x} - \underline{b}\|$ satisfies the normal equations*

$$A^\top A \underline{x}^* = A^\top \underline{b}.$$

Proof. Clearly, for every $\underline{x} \in \mathbb{R}^n$, $A\underline{x} \in \text{Col}(A)$. By Theorem 1.11.2 $\min_{\underline{x} \in \mathbb{R}^n} \|A\underline{x} - \underline{b}\| = \|A\underline{x}^* - \underline{b}\|$ if and only if $A\underline{x}^* = P_{\text{Col}(A)}\underline{b}$. But $\underline{b} - P_{\text{Col}(A)}\underline{b} \in \text{Col}(A)^\perp = \text{Null}(A^\top)$ by Theorem 1.10.10, and so,

$$\underline{0} = A^\top (\underline{b} - P_{\text{Col}(A)}\underline{b}) = A^\top (\underline{b} - A\underline{x}^*).$$

So $\underline{x}^*$ satisfies the equation $A^\top A \underline{x}^* = A^\top \underline{b}$. $\square$

Example 1.11.9. Find the solution of the linear equation $A\underline{x} = \underline{b}$ in the sense of the least squares and determine the error, where

$$A = \begin{bmatrix} 3 & 4 \\ -2 & -5 \\ 1 & -2 \end{bmatrix} \text{ and } \underline{b} = \begin{bmatrix} 10 & -6 & 3 \end{bmatrix}.$$

The least squares solution is given by:

$$\underline{x} = (A^\top A)^{-1} A^\top \underline{b}$$

Now

$$A^\top A = \begin{bmatrix} 3 & -2 & 1 \\ 4 & -5 & -2 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ -2 & -5 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 14 & 20 \\ 20 & 45 \end{bmatrix}$$

hence

$$(A^\top A)^{-1} = \frac{1}{230} \begin{bmatrix} 45 & -20 \\ -20 & 14 \end{bmatrix} = \begin{bmatrix} \frac{45}{230} & -\frac{20}{230} \\ -\frac{20}{230} & \frac{14}{230} \end{bmatrix} = \begin{bmatrix} \frac{9}{46} & -\frac{4}{23} \\ -\frac{4}{23} & \frac{7}{115} \end{bmatrix}$$

therefore

$$\underline{x} = (A^\top A)^{-1} A^\top \underline{b} = \begin{bmatrix} \frac{9}{46} & -\frac{4}{23} \\ -\frac{4}{23} & \frac{7}{115} \end{bmatrix} \begin{bmatrix} 3 & -2 & 1 \\ 4 & -5 & -2 \end{bmatrix} \begin{bmatrix} 10 \\ -6 \\ 3 \end{bmatrix} = \frac{1}{230} \begin{bmatrix} 745 \\ -4 \end{bmatrix}$$

The error vector is

$$\underline{r} = A\underline{x} - \underline{b} = \begin{bmatrix} 3 & 4 \\ -2 & -5 \\ 1 & -2 \end{bmatrix} \frac{1}{230} \begin{bmatrix} 745 \\ -4 \end{bmatrix} - \underline{b} = \begin{bmatrix} -\frac{81}{230} \\ -\frac{9}{23} \\ \frac{63}{230} \end{bmatrix},$$

hence the error is

$$\|\underline{r}\| = \sqrt{\left(\frac{81}{230}\right)^2 + \left(-\frac{9}{23}\right)^2 + \left(\frac{63}{230}\right)^2} = 0.44786829687102771022343 \dots$$

A particular example for that is when we want to fit a regression line to a data set.

Example 1.11.10 (Fitting a regression line). Let $(x_1, y_1), \dots, (x_n, y_n)$ be n points in $\mathbb{R}^2$. Find the constants a, b such that $y = ax + b$ approximates the n points best in sense of the least squared, i.e.

$$\sum_{i=1}^n (y_i - ax_i - b)^2$$

is minimal.

$$ax_i + b = y_i \text{ for every } i = 1, \dots, n \implies \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Let

$$A = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}.$$

Then

$$A^T A = \begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{bmatrix}, \text{ and } A^T \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix}.$$

Hence, solving the linear equation

$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix}.$$

leads us to

$$a = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2}, \quad b = \frac{(\sum_{i=1}^n x_i y_i)(\sum_{i=1}^n x_i) - n \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

and furthermore

$$\|Ay - b\| = \sqrt{\sum_{i=1}^n (y_i - ax_i - b)^2}.$$

Example 1.11.11. Let us consider the points $(-1, -2), (0, -3), (1, 2), (2, 3)$ on the plane. Using the method of the least squares, find the line which fits the best on these points. Determine the value of the error!

Computing

$$\begin{aligned} \sum x_i &= (-1) + 0 + 1 + 2 = 2, \\ \sum y_i &= (-2) + (-3) + 2 + 3 = 0, \\ \sum x_i^2 &= (-1)^2 + 0^2 + 1^2 + 2^2 = 1 + 0 + 1 + 4 = 6, \\ \sum x_i y_i &= (-1)(-2) + 0 \cdot (-3) + 1 \cdot 2 + 2 \cdot 3 = 2 + 0 + 2 + 6 = 10. \end{aligned}$$

yields

$$a = \frac{4 \cdot 10 - (2)(0)}{4 \cdot (6) - (2)^2} = 2, \quad b = \frac{(10)(2) - 4 \cdot 0}{4 \cdot 6 - (2)^2} = 1.$$

Hence the best fitting linear equation is $y = ax - b = 2x - 1$. The error is

$$r = \sqrt{(-2 - (2(-1) - 1))^2 + (-3 - (2(0) - 1))^2 + (2 - (2(1) - 1))^2 + (3 - (2(2) - 1))^2}$$

1.11.2 Exercises

1.11/1. Denote L the line spanned by the vector $\mathbf{a} = \begin{bmatrix} -6 \\ 8 \end{bmatrix}$ in $\mathbb{R}^2$.

(a) Give the matrix of the orthogonal projection from $\mathbb{R}^2$ to the line L (in the natural basis)!

(b) Find the orthogonal projection of the vector $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ to the line L !

1.11/2. Let us consider the vectors

$$\underline{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \underline{u}_2 = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} \text{ and } \underline{v} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}.$$

Let $V = \text{span}\{\underline{u}_1, \underline{u}_2\} \subset \mathbb{R}^3$ be the spanned vector space defined by the vectors $\underline{u}_1$ and $\underline{u}_2$.

(a) Find the matrix P_V of the orthogonal projection from $\mathbb{R}^3$ to V !

(b) Find the orthogonal projection of the vector $\underline{v}$ to the subspace V .

(c) Find the coordinates of the orthogonal projection of the vector $\underline{v}$ to the subspace V in the basis $B = \{\underline{u}_1, \underline{u}_2\}$!

1.11/3. Let $B = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}$. Show that there exists a subspace $V \subseteq \mathbb{R}^3$ such that B is the matrix of the orthogonal projection from $\mathbb{R}^3$ to V ! What is a basis of V ?

1.11/4. Find the solution with the method of the least squares of the equation $A\underline{x} = \underline{b}$, where

$$A = \begin{bmatrix} 2 & 1 \\ 4 & 2 \\ -2 & 1 \end{bmatrix} \text{ and } \underline{b} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}.$$

1.11/5. With the method of least squares, find the equation of the line which fits the best to the data set $(2, 1), (3, 2), (5, 3)$ and $(6, 4)$. Determine the error!

1.12 Decompositions of matrices

1.12.1 Definite matrices

Definition 1.12.1. Let $A \in \mathbb{R}^{n \times n}$ be an $n \times n$ symmetric matrix. We say that A is

- **Positive definite** if $\underline{x}^\top A \underline{x} > 0$ for all $\underline{x} \in \mathbb{R}^n \setminus \{\underline{0}\}$.
- **Positive semidefinite** if $\underline{x}^\top A \underline{x} \geq 0$ for all $\underline{x} \in \mathbb{R}^n$.
- **Negative definite** if $\underline{x}^\top A \underline{x} < 0$ for all $\underline{x} \in \mathbb{R}^n \setminus \{\underline{0}\}$.

- **Negative semidefinite** if $\underline{x}^\top A \underline{x} \leq 0$ for all $\underline{x} \in \mathbb{R}^n$.
- **Indefinite** if there exist $\underline{x}, \underline{y} \in \mathbb{R}^n$ such that $\underline{x}^\top A \underline{x} > 0$ and $\underline{y}^\top A \underline{y} < 0$.

Theorem 1.12.2. Let $A \in \mathbb{R}^{n \times n}$ be a $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. Then the followings hold:

- A is positive definite if and only if $\lambda_i > 0$ for all i .
- A is positive semidefinite if and only if $\lambda_i \geq 0$ for all i .
- A is negative definite if and only if $\lambda_i < 0$ for all i .
- A is negative semidefinite if and only if $\lambda_i \leq 0$ for all i .
- A is indefinite if and only if A has both positive and negative eigenvalues.

Proof. By Theorem 1.8.3, we can diagonalise A . That is, there exists an orthogonal matrix Q and

a diagonal matrix $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$ such that $A = QDQ^\top$. Then writing $\underline{y} = Q^\top \underline{x}$ we

have

$$\underline{x}^\top A \underline{x} = (Q^\top \underline{x})^\top D (Q^\top \underline{x}) = \underline{y}^\top D \underline{y} = \sum_{i=1}^n \lambda_i y_i^2.$$

So the sign of $\underline{x}^\top A \underline{x}$ for $\underline{x} \neq 0$ depends solely on the signs of the λ_i . For instance $\sum_{i=1}^n \lambda_i y_i^2 > 0$ for every $\underline{y} \neq 0$ then in particular for the choice $\underline{y}_i = \underline{e}_i$ and so $\lambda_i > 0$. And conversely, if $\lambda_i > 0$ for every i then for every $\underline{y} \neq 0$ there exists j such that $y_j \neq 0$, and so $\sum_{i=1}^n \lambda_i y_i^2 \geq \lambda_j y_j^2 > 0$. $\square$

Theorem 1.12.3. If $A \in \mathbb{R}^{n \times n}$ is symmetric and positive (semi)definite, then there exists a unique symmetric positive (semi)definite matrix B such that $B^2 = A$.

Proof. By Theorem 1.8.3, we can diagonalise A . That is, there exists an orthogonal matrix Q and

a diagonal matrix $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$ such that $A = QDQ^\top$. By Theorem 1.12.2, $\lambda_i \geq 0$

for every $i = 1, \dots, n$.

Define $\Lambda = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{\lambda_n} \end{bmatrix}$ and $B := Q\Lambda Q^\top$. Then B is symmetric and positive

semidefinite by Theorem 1.12.2. Furthermore,

$$B^2 = Q\Lambda Q^\top Q\Lambda Q^\top = Q\Lambda^2 Q^\top = A$$

since $Q^\top = Q^{-1}$ by the orthogonality of Q . $\square$

Example 1.12.4. Show that the matrix $A = \begin{bmatrix} 4 & -6 \\ -6 & 13 \end{bmatrix}$ is positive definite. Find a positive definite matrix B such that $B^2 = A$.

The characteristic polynomial of A is

$$\det(A - \lambda I) = (4 - \lambda)(13 - \lambda) - (-6)^2 = \lambda^2 - 17\lambda + 16,$$

Hence, the eigenvalues are

$$\lambda_{1,2} = \frac{17 \pm \sqrt{225}}{2} = \frac{17 \pm 15}{2} = \begin{cases} 16, \\ 1. \end{cases}$$

Both eigenvalues are positive, therefore A is positive definite.

Next compute an orthonormal eigenbasis. For $\lambda_1 = 16$ a corresponding eigenvector is

$$\underline{v}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad \|\underline{v}_1\| = \sqrt{1^2 + (-2)^2} = \sqrt{5}, \quad \underline{u}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Using Proposition 1.8.2, a corresponding eigenvector for $\lambda_2 = 1$ is then $\underline{u}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Let

$$Q = [\underline{u}_1 \quad \underline{u}_2] \quad \text{and} \quad D = \begin{bmatrix} 16 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then $A = Q\Lambda Q^\top$. A symmetric positive definite square root of A is

$$B = Q\Lambda Q^\top,$$

where $\Lambda = \begin{bmatrix} \sqrt{16} & 0 \\ 0 & \sqrt{1} \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$. After doing the matrix multiplication, we get

$$B = \frac{1}{5} \begin{bmatrix} 8 & -6 \\ -6 & 17 \end{bmatrix}.$$

1.12.2 Singular value decomposition (SVD)

?<thm:posdef2>?

Theorem 1.12.5. For any $A \in \mathbb{R}^{m \times n}$, the matrix $A^\top A$ is symmetric and positive semidefinite.

Proof. We have $\underline{x}^\top (A^\top A) \underline{x} = (A \underline{x})^\top (A \underline{x}) = \|A \underline{x}\|^2 \geq 0$. □

Definition 1.12.6. Let $A \in \mathbb{R}^{m \times n}$ be an $m \times n$ real matrix. The **singular values** of A are the non-negative square roots of the eigenvalues of $A^\top A$. They are usually denoted by $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n \geq 0$.

By Theorem 1.12.5, the matrix $A^\top A$ is symmetric and positive semidefinite, hence it has real, non-negative eigenvalues, so the singular values are well-defined.

?{thm:svd}?

Theorem 1.12.7 (Singular value decomposition (SVD)). *Let $A \in \mathbb{R}^{m \times n}$ be an $m \times n$ real matrix. There exist orthogonal matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ such that*

$$A = U\Sigma V^\top$$

where $\Sigma \in \mathbb{R}^{m \times n}$ is such that $\Sigma_{i,i} = \alpha_i$ for $i = 1, \dots, \min\{m, n\}$, and $\Sigma_{i,j} = 0$ for $i \neq j$, where α_i is the i th singular values of A .

Proof. We will only consider the proof of the special case when $m = n$ and A is invertible.

By Theorem 1.8.3, we can diagonalise the $n \times n$ symmetric matrix $A^\top A$. That is, there exist $n \times n$ orthogonal matrix $V = [\underline{v}_1 \ \cdots \ \underline{v}_n]$ with column vectors $\underline{v}_j$ and a diagonal matrix

$$\Lambda = \begin{bmatrix} \alpha_1^2 & 0 & \cdots & 0 \\ 0 & \alpha_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha_n^2 \end{bmatrix}, \text{ where } \alpha_i \text{ is the } i\text{th singular value of } A, \text{ and } A^\top A = V\Lambda V^\top. \text{ In particular,}$$

$$\Lambda = V^\top A^\top A V = (AV)^\top AV.$$

That is, $\Lambda_{i,j} = (A\underline{v}_i)^\top A\underline{v}_j$ and so, $A\underline{v}_j$ and $A\underline{v}_i$ are perpendicular if $i \neq j$, moreover, $\|A\underline{v}_i\|^2 = \alpha_i^2$ for every $i = 1, \dots, n$. By our assumption, A is invertible and so $\alpha_i > 0$ for every i .

Let us define the matrix $U = [\underline{u}_1 \ \cdots \ \underline{u}_n]$, where $\underline{u}_i = \frac{1}{\alpha_i} A\underline{v}_i$. Thus, U is also an orthogonal matrix. Furthermore, define $\Sigma = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha_n \end{bmatrix}$. Hence,

$$U = AV\Sigma^{-1} \Rightarrow U\Sigma V^\top = A,$$

which had to be proven. □

Note that if $m \neq n$ or A is not invertible then one can define U as follows: We may assume without loss of generality that $\alpha_1 \geq \cdots \alpha_k > 0 = \alpha_{k+1} = \cdots = \alpha_n$. Then let $U = [\underline{u}_1 \ \cdots \ \underline{u}_m]$ be such that $\underline{u}_i = \frac{1}{\alpha_i} A\underline{v}_i$ for $i = 1, \dots, k$ and by Gram-Schmidt orthogonalisation find an orthonormal basis $\{\underline{u}_{k+1}, \dots, \underline{u}_m\}$ for $\text{span}\{A\underline{v}_i : \alpha_i > 0\}^\perp$. Thus, U is also an orthogonal matrix.

?{ex:svd}?

Example 1.12.8. Find the singular value decomposition of $A = \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix}$.

The first step is to compute $A^\top A$ and its eigenvalues, (normalised) eigenvectors.

$$A^\top A = \begin{bmatrix} \sqrt{3} & 0 \\ 2 & \sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} = \begin{bmatrix} 3 & 2\sqrt{3} \\ 2\sqrt{3} & 7 \end{bmatrix}.$$

The characteristic polynomial is $\det(A^\top A - \lambda I) = (3 - \lambda)(7 - \lambda) - 12 = \lambda^2 - 10\lambda + 9$. Hence, the eigenvalues are $\lambda_1 = 9$, $\lambda_2 = 1$ and the singular values are $\alpha_1 = 3$, $\alpha_2 = 1$.

For $\lambda_1 = 9$:

$$(A^\top A - 9I)\underline{v}_1 = \underline{0} \Rightarrow \left[\begin{array}{cc|c} -6 & 2\sqrt{3} & 0 \\ 2\sqrt{3} & -2 & 0 \end{array} \right] \xrightarrow{\text{Gauss-Jordan elim.}} \left[\begin{array}{cc|c} 1 & -\frac{\sqrt{3}}{3} & 0 \\ 0 & 0 & 0 \end{array} \right] \Rightarrow \underline{v}_1 = \begin{bmatrix} \frac{\sqrt{3}}{3}y \\ y \end{bmatrix}.$$

Normalising the vector $\underline{v}_1$ we get $\frac{1}{3}y^2 + y^2 = 1$. We get that $y = \pm \frac{\sqrt{3}}{2}$ and so, $\underline{v}_1 = \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$.

We know that the eigenvectors of $A^\top A$ are pairwise orthogonal so we might choose $\underline{v}_2 = \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}$, as the normalised eigenvector for the eigenvalue $\lambda_2 = 1$. Thus,

$$V = [\underline{v}_1 \quad \underline{v}_2] = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}.$$

Finally, we calculate the column vectors of U via $\underline{u}_i = \frac{1}{\alpha_i} A \underline{v}_i$. For $i = 1$:

$$\underline{u}_1 = \frac{1}{3} A \underline{v}_1 = \frac{1}{3} \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}.$$

Similarly, for $i = 2$:

$$\underline{u}_2 = \frac{1}{\alpha_2} A \underline{v}_2 = \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}.$$

Thus,

$$U = [\underline{u}_1 \quad \underline{u}_2] = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

The singular value decomposition has an important geometric interpretation. Namely, it describes the image of the unit ball under the linear transformation $\underline{x} \mapsto A\underline{x}$, see Figure 1.5.

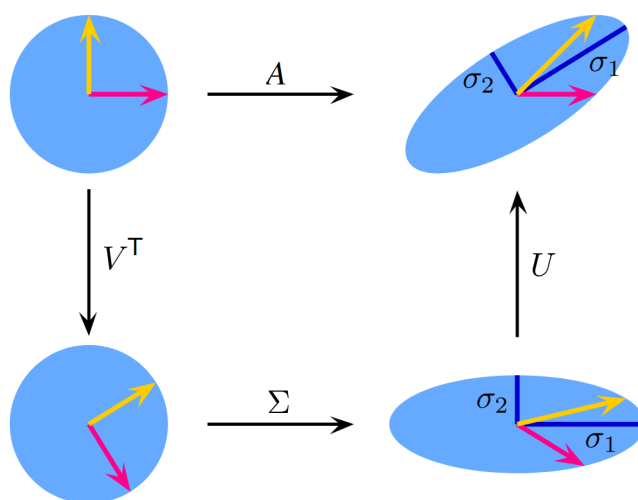


Figure 1.5: The geometric meaning of singular value decomposition: In Example 1.12.8, V is a rotation by 60° , Σ scales by factors 3 and 1, and U is a rotation by 30° . So A maps the unit circle to an ellipse. The matrix $A = U\Sigma V^\top$ maps the unit circle to an ellipse whose principal axes are along the columns of U with semi-axis of length $\sigma_1 = 3$ and $\sigma_2 = 1$.

?<fig:SVD>?

prop:polardec)?

Proposition 1.12.9 (Polar decomposition). *For a $n \times n$ real matrix $A \in \mathbb{R}^{n \times n}$, there exist a symmetric positive semidefinite $n \times n$ matrix P and an orthogonal Q such that*

$$A = PQ$$

If A is invertible then P is positive definite and Q is unique.

Proof. By Theorem 1.12.7, $A = U\Sigma V^\top$, where U and V are $n \times n$ orthogonal matrices and Σ is the diagonal matrix formed by the singular values $\alpha_1 \geq \dots \geq \alpha_n \geq 0$. Consider the matrix $P = U\Sigma U^\top$. By Theorem 1.12.2, P is symmetric positive semi definite and if A is invertible then $\alpha_n > 0$ and so, P is positive definite. Moreover, $A = U\Sigma V^\top = U\Sigma U^\top UV^\top = PQ$, where $Q = UV^\top$ is an orthogonal matrix. $\square$

Example 1.12.10. Find the polar decomposition of $A = \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix}$.

From Exercise 1.12.8, we get that the singular value decomposition of A is

$$A = U\Sigma V^\top, \text{ where } U = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{-1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}, \Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } V = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}.$$

By the construction in the proof of Proposition 1.12.9

$$P = U\Sigma U^\top, \quad Q = UV^\top.$$

Computing P and Q by using matrix multiplication, we get

$$P = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{-1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{-1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} \frac{5}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{3}{2} \end{bmatrix}.$$

Now compute

$$Q = UV^\top = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix},$$

which is orthogonal, as the decomposition required it to be.

Let us observe that P is the unique positive semi-definite symmetric matrix such that $P^2 = AA^\top$, which might provide us an alternative construction for P .

1.12.3 Spectral decomposition

Proposition 1.12.11 (Spectral decomposition). *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric $n \times n$ real matrix. Then*

$$A = \sum_{i=1}^n \lambda_i \underline{u}_i \underline{u}_i^\top,$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues and $\underline{u}_1, \dots, \underline{u}_n$ are the corresponding eigenvectors of A forming an orthonormal basis of $\mathbb{R}^n$.

Proof. By Theorem 1.8.3, $A = QDQ^\top$, where $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$ and $Q = [\underline{u}_1 \ \underline{u}_2 \ \cdots \ \underline{u}_n]$.

Then

$$A = QDQ^\top = \begin{bmatrix} \lambda_1 \underline{u}_1 & \lambda_2 \underline{u}_2 & \cdots & \lambda_n \underline{u}_n \end{bmatrix} \begin{bmatrix} \underline{u}_1^\top \\ \underline{u}_2^\top \\ \vdots \\ \underline{u}_n^\top \end{bmatrix} = \sum_{i=1}^n \begin{bmatrix} 0 & \cdots & 0 & \lambda_i \underline{u}_i & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \underline{u}_1^\top \\ \underline{u}_2^\top \\ \vdots \\ \underline{u}_n^\top \end{bmatrix} = \sum_{i=1}^n \lambda_i \underline{u}_i \underline{u}_i^\top$$

by the basic law of matrix multiplication. $\square$

Note that the matrix $\underline{u}_i \underline{u}_i^\top$ is the matrix of orthogonal projection to the subspace $\text{span}\{\underline{u}_i\}$. In particular, every symmetric $n \times n$ real matrix can be written as a linear combination of orthogonal projections. Moreover,

$$A^2 = \sum_{i=1}^n \lambda_i \underline{u}_i \underline{u}_i^\top \sum_{j=1}^n \lambda_j \underline{u}_j \underline{u}_j^\top = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \underline{u}_i \underline{u}_i^\top \underline{u}_j \underline{u}_j^\top = \sum_{i=1}^n \lambda_i^2 \underline{u}_i \underline{u}_i^\top,$$

where we used that $\underline{u}_i^\top \underline{u}_j = 0$ if $i \neq j$. In particular, for any $k \geq 1$

$$A^k = \sum_{i=1}^n \lambda_i^k \underline{u}_i \underline{u}_i^\top. \quad (1.12.1) \text{ ?eq:power?}$$

Definition 1.12.12 (Matrix valued functions). *Let A be an $n \times n$ symmetric real matrix with eigenvalues $\lambda_1, \dots, \lambda_n$. Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a real analytic function on an interval containing $\{\lambda_1, \dots, \lambda_n\}$. Then we define*

$$f(A) := \sum_{i=1}^n f(\lambda_i) \underline{u}_i \underline{u}_i^\top,$$

where $\sum_{i=1}^n \lambda_i \underline{u}_i \underline{u}_i^\top$ is the spectral decomposition of A .

The definition is coherent in the following sense: Since $f: \mathbb{R} \rightarrow \mathbb{R}$ is analytic, we can write $f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$, where $f^{(k)}$ denotes the k th derivative of f . By this analogy, we define

$$f(A) := \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} A^k.$$

But applying (1.12.1) to calculate the matrix powers, and so

$$f(A) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} A^k = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} \sum_{i=1}^n \lambda_i^k \underline{u}_i \underline{u}_i^\top = \sum_{i=1}^n \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} \lambda_i^k \underline{u}_i \underline{u}_i^\top = \sum_{i=1}^n f(\lambda_i) \underline{u}_i \underline{u}_i^\top.$$

Example 1.12.13. Consider the symmetric matrix

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

(a) Find the spectral decomposition of A .

(b) Find the matrix e^{2A} .

Compute the eigenvalues of A from $\det(A - \lambda I) = 0$:

$$\det \begin{bmatrix} 1 - \lambda & 2 \\ 2 & -2 - \lambda \end{bmatrix} = (1 - \lambda)(-2 - \lambda) - 4 = \lambda^2 + \lambda - 6 = 0.$$

Hence, the eigenvalues are

$$\lambda_{1,2} = \frac{-1 \pm \sqrt{1 - 4(-6)}}{2} = \begin{cases} 2 \\ -3. \end{cases}$$

For $\lambda_1 = 2$ solve $(A - 2I)\underline{v} = \underline{0}$:

$$\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \underline{0} \Rightarrow -x + 2y = 0 \Rightarrow x = 2y.$$

Take $\underline{v}_1 = \begin{bmatrix} 2y \\ y \end{bmatrix}$, normalizing it we get that $(2y)^2 + y^2 = 1$ and so $y = \frac{\pm 1}{\sqrt{5}}$. In particular,

$$\underline{u}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Using that the eigenvectors of a symmetric matrix are pairwise orthogonal (Proposition 1.8.2), we can choose the eigenvector of the eigenvalue $\lambda_2 = -3$ to be

$$\underline{u}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Hence, the spectral decomposition theorem gives

$$A = \lambda_1 \underline{u}_1 \underline{u}_1^\top + \lambda_2 \underline{u}_2 \underline{u}_2^\top = 2 \underline{u}_1 \underline{u}_1^\top - 3 \underline{u}_2 \underline{u}_2^\top. \quad (1.12.2) \text{ ?eq:specdec?}$$

Explicitly,

$$\underline{u}_1 \underline{u}_1^\top = \begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix}, \quad \underline{u}_2 \underline{u}_2^\top = \begin{bmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{bmatrix},$$

so one may check

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

as required.

(b) To compute e^{2A} , we use the spectral decomposition in (1.12.2). Thus,

$$e^{2A} = e^4 \underline{u}_1 \underline{u}_1^\top + e^{-6} \underline{u}_2 \underline{u}_2^\top.$$

Using the matrices for $\underline{u}_1 \underline{u}_1^\top$ and $\underline{u}_2 \underline{u}_2^\top$ above,

$$e^{2A} = e^4 \begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix} + e^{-6} \begin{bmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{bmatrix} = \begin{bmatrix} \frac{4e^4 + e^{-6}}{5} & \frac{2e^4 - 2e^{-6}}{5} \\ \frac{2e^4 - 2e^{-6}}{5} & \frac{e^4 + 4e^{-6}}{5} \end{bmatrix}.$$

1.12.4 Exercises

1.12/1. Let

$$A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$$

(a) Show that the matrix A is positive definite.

(b) Find a positive definite matrix B such that $B^2 = A$!

1.12/2. Find the singular value decomposition of the matrix $A = \begin{bmatrix} 3 & 8 \\ -4 & 6 \end{bmatrix}$.

1.12/3. Find the singular value decomposition of the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$.

1.12/4. Consider the matrix $A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$.

(a) Give the spectral decomposition of the matrix A .

(b) Find the matrix e^A .

1.13 Mixed exercises in Linear Algebra

1.13/1. Consider the matrix $A = \begin{bmatrix} 4 & 5 & 6 \\ 3 & -1 & 2 \\ 0 & 2 & 1 \end{bmatrix}$.

(a) Find the determinant $\det(A)$.

(b) If it is invertible find the inverse matrix A^{-1} .

1.13/2. Solve the following linear equation system:

$$\begin{aligned}x_1 + 3x_2 - 2x_3 + 2x_5 &= 0 \\2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 &= -1 \\5x_3 + 10x_4 + 15x_6 &= 5 \\2x_1 + 6x_2 + 8x_4 + 4x_5 + 18x_6 &= 6.\end{aligned}$$

1.13/3. Is the collection of vectors linearly independent?

$$\underline{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \underline{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \text{ and } \underline{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix}.$$

If not then choose a maximal linearly independent subset and express the remaining vector as a linear combination of these vectors.

1.13/4. Consider two bases $\mathcal{B} = \left\{ \underline{u}_1 = \begin{bmatrix} 15 \\ -2 \end{bmatrix}, \underline{u}_2 = \begin{bmatrix} 8 \\ -1 \end{bmatrix} \right\}$ and $\mathcal{B}' = \left\{ \underline{v}_1 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \underline{v}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ of $\mathbb{R}^2$.

(a) Find the basis transformation matrices $P_{\mathcal{N} \rightarrow \mathcal{B}}$ and $P_{\mathcal{B}' \rightarrow \mathcal{B}}$.

(b) If $[\underline{w}]_{\mathcal{N}} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ then what is $[\underline{w}]_{\mathcal{B}} = ?$

(c) If $[\underline{z}]_{\mathcal{B}'} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ then what is $[\underline{z}]_{\mathcal{B}} = ?$

1.13/5. A linear transformation T maps the vectors $\mathcal{B} = \left\{ \underline{u}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \underline{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \underline{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ to the

vectors $\underline{v}_1 = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}, \underline{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\underline{v}_3 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$ respectively (that is $T(\underline{u}_i) = \underline{v}_i$ for every $i = 1, 2, 3$).

(a) Find the matrix representation $[T]_{\mathcal{B}}$ of the linear transformation T in the basis $\mathcal{B}$.

(b) Find the matrix representation $[T]_{\mathcal{N}}$ of T in the natural basis $\mathcal{N}$.

1.13/6. Find the eigenvalues and eigenvectors and diagonalise the matrix $A = \begin{bmatrix} 2 & 1 \\ 4 & 1 \end{bmatrix}$.

1.13/7. Determine an orthonormal basis for the subspace spanned by the vectors

$$\underline{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 1 \end{bmatrix}, \underline{u}_2 = \begin{bmatrix} 4 \\ 6 \\ 5 \\ 1 \end{bmatrix} \text{ and } \underline{u}_3 = \begin{bmatrix} 5 \\ 4 \\ -2 \\ 0 \end{bmatrix}.$$

1.13/8. Draw the points on the plane, which satisfy the equation $5x^2 - 4xy + 8y^2 = 36$.

1.13/9. Let

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

(a) Give a basis of the column space $\text{Col}(A)$ out of the column vectors of A , and express the coordinates of the remaining vectors in this basis.

(b) What are $\text{rank}(A)$, $\text{rank}(A^T A)$, $\text{nullity}(A^T)$ and $\text{nullity}(A)$?

1.13/10. Determine the matrix of the orthogonal projection (in the natural basis) from $\mathbb{R}^2$ to the line $y = \frac{\sqrt{3}}{2}x$. What is the orthogonal projection of the point $\underline{v} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$?

1.13/11. Determine the matrix of the orthogonal projection (in the natural basis) from $\mathbb{R}^3$ to the plane $\left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : 5x - 6y + z = 0 \right\}$. What is the orthogonal projection of the point $\underline{v} = \begin{bmatrix} 8 \\ 1 \\ 3 \end{bmatrix}$?

1.13/12. Find the equation of the line, which fits the best (in the sense of least squares) to the points $(-3, -2)$, $(0, 3)$, $(1, 1)$ and $(2, 0)$, and determine the error!

1.13/13. Let $A = \begin{bmatrix} 9 & 6 \\ 6 & 9 \end{bmatrix}$

(a) Show that A is positive definite.

(b) Find a positive definite symmetric matrix B such that $B^2 = A$.

1.13/14. Let $A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$.

(a) Find the singular value decomposition of the matrix A .

(b) Find the polar decomposition of the matrix A .

1.13/15. Let $A = \begin{bmatrix} -13 & 20 \\ -\frac{15}{2} & 12 \end{bmatrix}$.

(a) Find the spectral decomposition of the matrix A .

(b) Determine the matrix $\cos(\pi A)$.